

FAKULTA
INFORMAČNÍCH
TECHNOLOGIÍ
ÚSTAV
INFORMAČNÍCH
SYSTÉMŮ

Pokročilé databázové systémy
PDB

Multimediální databáze

Version 0.0.0.9
Evaluation.

Poděkování

Děkuji všem, kteří mě při psaní této práce podporovali, především J. Zendulkovi a A. Ženíškové.

Děkuji také za všechny připomínky.

Organizace studijní opory

Odhadovaný čas potřebný pro studium celého materiálu:

- ⌚ **16 hodin +-8 hodin.**

To jsou asi 2 pracovní dny (+-1den).

Následuje odhadovaný čas rozepsaný po kapitolách včetně jejich popisu.

První kapitola je věnována letnému úvodu a historii multimediálních databází,

- ⌚ 1 hodina.

Ve druhé kapitole jsou uvedeny základní pojmy multimediálních databází a požadované vlastnosti systému řízení báze multimediálních dat, včetně jejich vysvětlení. Dále popis standardů MPEG pro manipulaci s multimediálními daty a popis jejich obsahu,

- ⌚ 8 hodin.

Třetí část je věnována standardům potřebným k dobrému porozumění současným multimediálním databázovým systémům,

- ⌚ 5 hodin.

(tato část není oporou přednášky doc. Zendluky na téma objektově-relační databáze)

Čtvrtá část je věnována komerčním systémům, především Oracle® *interMedia*,

- ⌚ 1 hodina.

Závěr je věnován shrnutí a otevřeným otázkám multimediálních databází,

- ⌚ 1 hodina.

Součástí této práce jsou přílohy ve formě prezentace. První z nich stručně prezentuje základní poznatky této práce; druhá slouží jako jednoduchá příručka praktické realizace multimediální databáze v Oracle® *interMedia*.

Obsah

1	Úvod.....	4
1.1	Historie	4
1.2	Normalizace.....	5
2	Multimediální databázové systémy.....	6
2.1	Databázová diverzita	6
2.2	Multimediální databáze	6
2.3	Požadavky na konstrukci MMDBMS	9
2.4	Multimédia.....	11
2.5	Formáty audiovizuálního obsahu	12
2.5.1	Obrázky	12
2.5.2	Dokumenty	12
2.5.3	Zvuk.....	12
2.5.4	Video a mediální kontejnery	13
2.6	Metadata	14
2.6.1	MPEG-7.....	15
2.6.2	MPEG-21.....	18
2.7	Extrakce rysů a popis obsahu	18
2.7.1	Popis obsahu audia	18
2.7.2	Extrakce rysů z obrázků	19
2.7.3	Tvar	20
2.7.4	Pohyb a umístění	21
2.7.5	Speciální aplikace.....	22
2.8	Vyhledávání.....	23
2.8.1	Podobnostní vyhledávání	24
2.9	Indexace.....	25
2.9.1	K-D stromy.....	26
2.9.2	Quad stromy	26
2.9.3	R-stromy	27
2.10	Ukládání.....	27
2.11	Přenos	28
2.12	Prezentace	30
2.13	Bezpečnost a ochrana	30
2.14	Architektury multimediálních systémů	30
2.15	Aplikace získávání znalostí	31
2.15.1	Klasifikace.....	31
2.15.2	Shlukování a analýza.....	34
3	Databázové technologie a standardy.....	37
3.1	Objektově orientovaný koncept.....	37

3.1.1	Objektový model ODMG.....	38
3.1.2	Definiční jazyk ODL.....	39
3.1.3	Dotazovací jazyk OQL.....	41
3.2	Objektově relační koncept.....	42
3.2.1	SQL:1999.....	42
3.2.2	SQL:2003.....	44
3.3	SQL / MM.....	45
4	Komerční databázové systémy.....	47
4.1	Oracle interMedia.....	47
5	Shrnutí.....	50

1 Úvod

Požadavek multimediálních databází vznikl postupně v mnoha odvětvích. Obchodní služby a průmysl (CAD, online prodej, video a audio na přání) potřebují skladovat a distribuovat velké množství multimediálních dat. Vzdělávání (e-learning), státní správa (bezpečnost) a zdravotnictví vyžadují více. Především je to možnost dotazovat multimediální data na základě jejich obsahu.

Multimediální systémy řízení báze dat (MMDBMS¹) poskytují podporu pro správu multimediálních dat jako doplněk tradičních databázových systémů. Proto musí obsahovat obdobné prostředky pro definici multimediálních typů dat, jejich modelování, skladování, manipulaci, dotazování, řízení přístupu, organizaci a prezentaci nezávislou na typu média, i když jejich hlavním úkolem je obvykle efektivně poskytovat multimediální data a informace, které jsou v multimediálních datech obsaženy.

Pod multimediálními (MM) daty si lze představit nestrukturovaná data, která dělíme na základě způsobu percepcie informace na vizuální a audio. Mezi audio patří hudba, řeč a ostatní zvuky. Vizuální informace jsou například statické obrázky, ty chápeme jako 2D grafiku (vektorová, obchodní) a další obrázky (foto), které někdy lze reprezentovat pomocí 3D modelu. Video je sekvence (pohyblivých) obrázků, volitelně doplněná o zvuk nebo text. Textová informace může představovat vlastní kategorii, nebo ji spolu s formátovanými dokumenty doplněnými o obrázky řadíme mezi vizuální data.

Multimediální databáze (MMDBs) a jejich aplikace mají v poslední době výrazně rostoucí trend. To je dáno třemi předpoklady – rozšířením počítačů, růstem jejich výpočetní síly a paměťové kapacity, rychlostí komunikačních technologií a nízkou cenou záznamových, zobrazovacích a přehrávacích zařízení (digitální fotoaparáty, audio a video přehrávače).

1.1 Historie

Počátek multimediálních databází lze stěží specifikovat přesně. Pokud bychom je chápali na určité úrovni abstrakce jako kolekci vizuálních dat, která je nějakým způsobem organizována, mohly by jejich počátky sahát až do starověkého Egypta. V případě dalších multimediálních dat do doby vynálezu záznamu zvuku a videa, konce 19. století.

S jistotou lze říci, že multimediální databáze založené na počítačích neexistovaly dříve, než byly vytvořeny první systémy pro ovládání souborů a řízení báze dat v 60. letech. Tehdy byly stanoveny také objektově orientované přístupy (Simula). Relační model dat byl představen začátkem 70. let. Organizované skladování multimediálních dat ale nebylo prakticky realizovatelné až do konce 80. let, počátku prvních objektově orientovaných systémů řízení báze dat (Graphael, Servio, Ontos). Byla to multimediální data spolu s prostorovými, která vytvořila požadavek na sblížení relačního a objektového přístupu začátkem 90. let (UniSQL, Postgres). Perzistentní existenci objektů požadovalo mnoho výrobních odvětví (simulace, CAD, CAM) nebo medicína, kartografie a také věda, zejména fyzika, chemie a biologie.

První multimediální databázové systémy začaly vznikat koncem 80. let (ORION, JASMIN, MediaWay – viz Chuckwalla, kapitola 0). Jejich vývoj se výrazně urychlil v 90. letech spolu s boomem multimedií, Javy, XML a internetu na osobních počítačích

¹ MultiMedia DataBase Management System

následovaný různými ad-hoc zařízeními (digitální TV, mobilní telefon, iPod) na začátku 21. století s trvalým požadavkem na multimediální komunikaci a zábavu.

1.2 Normalizace

Těžko by si mohl někdo v Japonsku poslechnout MP3 (MPEG1 Audio layer 3) skladbu Francouzské zpěvačky, nahranou v Anglii, uloženou na serveru v USA pomocí Finského telefonu, nebýt dostatečné spolupráce mezinárodních normalizačních institucí – International Organization for Standardization (ISO), International Electronic Commision (IEC) a International Telecommunications Union (ITU).

Společnosti a instituce, mající nějakou roli v informačních technologiích, si uvědomily, že různé inovace, vývoj novějších a kvalitnějších produktů, případně přenos aplikace na výhodnější platformu se nemusí vyplatit vzhledem k nutné investici. Proto vytvořily v průběhu 80. let konsorcia s cílem standardizovat určitou oblast IT.

Příkladem této snahy je dotazovací jazyk pro relační databáze SQL (ISO 1987), jehož standard se průběžně vyvíjí (ISO 1992, 1999, 2003) a rozšiřuje – o správu objektů, XML a funkce OLAP, nebo o multimédia a aplikační balíčky (SQL/MM, ISO 2003).

Ne vždy proces normalizace nabude schválené mezinárodní normy. Příkladem je zmíněný SQL/MM, jehož vývoj se zastavil u statických obrázků. Nebo také první standard popisující perzistentní objekty v objektových databázích skupiny Object Data Management Group (ODMG, 1993-2000). Ale jejich spojení se skupinou Object Management Group (OMG) vyústilo ve všeobecně uznávaný standard Unified Modeling Language (UML, ISO 2005).

V multimediální oblasti jsou nejznámější skupiny Joint Photographic Experts Group (JPEG, ISO 1994, 1999 a JPEG 2000) a Motion Picture Expert Group (MPEG-1, ISO 1993, MPEG-2 1999 a MPEG-4 2004 pro kompresi videa, MPEG-7 2005 pro popis MM obsahu a MPEG-21 2007 pro jejich distribuci).

Mezinárodních norem existuje velké množství, další jsou například z oblasti komunikací (zmíněné dále v textu), nebo Slovník informačních technologií (ISO: 1999). Mimo ně jsou uznávány i jiné standardy, například otevřené (RFC HTTP, IP) nebo firemní standardy (Adobe PDF, MS Office).

Zamyšlení k 1. kapitole

Úvod neposkytuje příliš prostoru pro doplňující otázky, slouží pouze k seznámení se se základními pojmy a zkratkami a zjištění, že nezávislý výzkum se sjednocuje prostřednictvím norem.

Jak a proč multimediální aplikace vznikly a k čemu se využívají?

Více o multimediálních datech a databázích se dozvíte v následující kapitole.

2 Multimediální databázové systémy

Velké množství dat, různých (i multimediálních) aplikací, pracujících s rozmanitými daty jednoduše požaduje databáze, protože databáze zaručují konzistenci, souběžnost, integritu, bezpečnost a dostupnost dat. Z perspektivy uživatele poskytují databáze funkce pro jednoduchou manipulaci, dotazování a získávání vysoce relevantních informací z obrovských kolekcí produkovaných a uložených dat.

2.1 Databázová diverzita

Podíváme se na databáze z hlediska jejich diverzity – typu dat, která obsahují. Předem je jednoduše rozdělíme na strukturovaná a nestrukturovaná data.

Strukturovaná data – klasická, nejčastěji časově uspořádaná (temporální) data – text a číselné informace s časovými razítky vhodné pro informační systémy obchodních společností a institucí. Často také mívají transakční charakter, například nákup nebo platební operace. Podle způsobu použití těchto databází ještě rozlišujeme:

- Transakční zpracování (OLTP²) je charakteristické množstvím uživatelů a operací pro manipulaci (insert, update, delete) a vyhledávání.
- Datové sklady a OLAP³ obsahují velké objemy (strukturovaných) dat, určených primárně pro jejich analýzu a podporu rozhodování (dolování dat).

Pro strukturovaná data byly navrženy relační databáze. Tyto databáze mohou mít volitelně objektové rozšíření, případně obsahovat odkazy na externí zdroje. Pro nestrukturovaná data jsou tyto postrelační (objektově orientované) vlastnosti nezbytné.

Nestrukturovaná data jsou speciální data, jejichž struktura není známa, nebo je složitá. Jejich příkladem jsou různé fyzikální modely částic nebo vesmíru, chemické databáze molekul, medicínské záznamy DNA a podobně. Do nestrukturovaných, respektive složitě strukturovaných dat počítáme i následující obvyklé typy dat:

- **Textová data** – data s velmi složitou strukturou. Jejich příkladem je beletrie.
- **Strukturovaný text** zahrnuje strukturované (formátované) dokumenty, webové stránky, XML dokumenty nebo zdrojové kódy programů.
- **Prostorová data** – klasická data doplněná o jejich umístění, tvar a rozměry.
- **Multimediální data** – digitalizovaná audiovizuální data (viz kapitola 2.4).

Tato data se v databázích nevyskytují samostatně. Systémy řízení báze složitě strukturovaných dat poskytují podporu pro správu těchto dat jako doplněk tradičních databázových systémů. Proto musí obsahovat obdobné prostředky pro definici nestrukturovaných typů dat, jejich modelování, skladování, manipulaci, dotazování, řízení přístupu, organizaci a prezentaci nezávislou na typu, i když jejich hlavním úkolem je obvykle efektivně poskytovat požadovaná data a přidružené informace.

2.2 Multimediální databáze

Databáze není jedinou možností, jak mít data organizována. Stejně je to i s multimediálními daty. Pokud si archivujete video z dovolené na kazetách nebo DVD a vyhovuje vám to, není důvod měnit. Pokud ale máte médií velké množství (jako

² Online Transaction Processing (OLTP)

³ Online Analytical Processing (OLAP)

televize), není jejich uchovávání (ve sklepech), správa (těžká), vyhledávání (vyčerpávající) a doručování (vlakem) příliš efektivní.

Tento problém je částečně řešitelný běžnou počítačovou technikou – pomocí diskového pole, případně zveřejněné prostřednictvím WWW nebo FTP serveru. Výsledek bude uspokojivý v případě textu nebo obrázků, ale těžko v případě (real-time) proudů audia a videa. Hlavním důvodem je omezená přenosová kapacita. Proto je výhodné použít tzv. streaming server, který zajišťuje efektivní využití přenosového pásma podle požadovaného média na přání, případně (multicast) živého vysílání a některé další funkce převzaté z databázového prostředí.

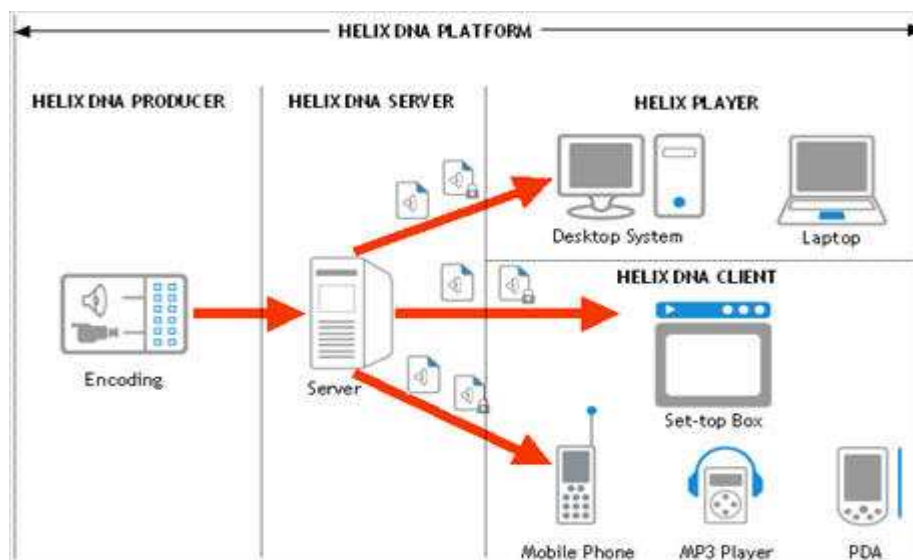


Fig. 2.1 Architektura streaming serveru (převzato z www.helixcommunity.org)

Takovýmto (Fig. 2.1) způsobem je možné provozovat televizní vysílání, video a audio na přání a podobně. Příkladem je Internetové vysílání ČT (www.ceskatelevize.cz/vysilani/), Českého rozhlasu (www.rozhlas.cz/audio/vysilani/), nebo záznamy a streaming přednášek Video serverem FIT (video1.fit.vutbr.cz) nebo množství komerčních a komunitních poskytovatelů multimediálního obsahu.

Popsané řešení se nicméně stává problematickým pro vyhledávání. Vyhledávání je stěžejní problém multimediálních dat. Vyhledání čísla, slova, případně názvu souboru je mnohem jednodušeji představitelné a realizovatelné, než vyhledání požadovaného objektu na obrázku, akce ve videu, slova nebo melodie ve zvukových datech. Vyhledávání v multimediálních datech se tedy děje výhradně pomocí metadat. Metadata jsou data o datech, jejich popis.

Říká se, že jeden obrázek vydá za sto slov. Pro popis obrázku k vyhledávání jich stačí mnohem méně. Problémem ale je, že se subjektivně liší, člověk od člověka (bracha.jpg, 8308141234.jpg). Jeho práce je navíc pro tento účel příliš drahá a neefektivní. Požadujeme tedy objektivní, automatické metody pro strukturovaný popis obsahu multimedií. Takovýto deskriptor⁴ lze získat extrakcí charakteristických rysů z média. Jak, je blíže uvedeno v kapitole 2.7.

Na základě způsobu popisu rozlišujeme dva základní typy vyhledávání:

- **Vyhledávání dle popisu dat** (description-based retrieval). Jedná se o deskripci sémantiky dat. Někakým způsobem je nutné přiřadit název, zapsat klíčová slova, pojmenování osob, objektů, akcí, zvuků – a to nelze obecně provést žádným

⁴ Také vektor rysů nebo signatura.

automatizovaným systémem. Popis musí vytvořit člověk při jejich tvorbě (například na webu). Následný systematický popis při procházení a třídění člověkem není vůbec jednoduchý, rychlý ani levný způsob, nehledě na neúplnost a inkonzistenci způsobenou subjektivním přístupem, závislým na kontextu.

- **Vyhledávání dle obsahu dat** (content-based retrieval, CBR). Jedná se o metodu popisu syntaxe, obvykle plně automatizovatelnou. To bohatě nahrazuje její nedostatky, zejména co se týká kvality popisu sémantiky. V případě zvukových dat se může jednat o rozpoznání klíčových slov, nástrojů, not, tempa nebo i žánru. Při popisu obrazových dat, máme k dispozici dominantní barvu, histogram, hrany, textury, tvary, regiony, umístění, pohyb, popis obličeje nebo jiné biometrické informace.

Multimediální aplikace vyhledávající dle popisu dat se používají především v zábavním průmyslu pro vyhledání filmu s hledaným hercem, skladby jistého interpreta, žánru a podobně. Známé jsou Exif a ID3 popisy (kapitola 2.6). V informačním průmyslu se používá pro anotace fotografií, zvukových nahrávek a přepis mluveného slova do titulků. V průmyslu slouží například pro popis CAD/CAM výkresů, různé aplikace zasahují snad do všech odvětví. V kartografii a dalších oborech se snaží oprostít od manuální anotace snímků vzdáleného pozorování (www.mapy.cz) a používají automatickou detekci požadovaných prostorových objektů, které poté zkontrolují, případně ručně upřesní. Zajímavým projektem je NASA Earth Observing System (eosps.gsfc.nasa.gov).

Kombinace manuálního a automatického vyhledávání je dobře použitelná v případě, že známe útržky z názvu snímku, datum pořízení, nebo klíčová slova a z jeho obsahu mohou být extrahovány další vlastnosti, například diagnóza.

Vyhledávání zaměřené na obsah má v reálném světě stále více aplikací. Důležité je v medicíně, například pro detekci anomálních tkání v různých typech snímků (kapitola 2.8), v meteorologii pro zmapování a následnou předpověď počasí. Vyhledávače obrázků na Internetu dle obsahu jsou v současné době ve stádiu vývoje, speciální aplikace ale existují v mnoha průmyslových a vědních odvětvích. Například systémy pro nalezení nových vesmírných těles fungují spolehlivě. Stejně tak rozpoznání textu (Optical Character Recognition, OCR), za předpokladu rozumné kvality média.

V případě mluveného slova – jeho přepis, identifikace mluvčího, jazyka a podobně je na světové špičce skupina Speech@FIT. V Hudebně například [Fraunhofer IDMT](#).

Stále důležitější jsou také bezpečnostní aplikace. V současné době jsou schopny identifikovat člověka podle obličeje, oční duhovky, otisku prstu a dalších antropometrických vlastností. Často mají za úkol sledovat pohybující se objekty a analyzovat jejich počet, stav a chování. Ať už jde o osoby, nebo automobily – jízda na červenou, překročení rychlosti, nedovolené předjíždění, případně odcizená vozidla nebo značky včetně identifikace pachatele za volantem.

Popis na základě obsahu využívá také vlastností lidského vnímání. Zejména pro nalezení podobnosti, což je v mnoha aplikacích žádoucí. Ve vyhledávání zaměřeném na obsah se obvykle používají dva typy dotazů:

- **Předložení vzorového obrazu** (similarity search). Využívá se při tom porovnání deskriptorů extrahovaných z obrázku dotazu s deskriptory obrazů, které byly předem extrahovány a uloženy do databáze. Výsledek dotazu je například několik obrázků z databáze, jejichž deskriptory jsou dotazovanému nejbližší, a tedy by měly být tomu dotazovanému nejpodobnější.

- **Specifikace vlastností** – prvků, ze kterých je tvořen deskriptor. Jsou porovnány s těmi v databázi. Můžeme například zadat klíčové slovo, barvu, tvar (náčrtkem), specifikovat texturu, nebo parametry melodie (zanotovat).



Fig. 2.2 Dotaz pro podobnostní vyhledávání a specifikace vlastností Mony Lizy

Procesy v multimediální databázi lze zjednodušeně pochopit z ilustrace Fig. 2.3. Před uložením MM dat do databáze je provedena extrakce rysů (Feature extraction), které tvoří deskriptor (AV Description) a jsou společně uložena do databáze (Storage). Uživatelé a jiné služby mohou tuto databázi dotazovat (Search / query), výsledky dotazů prohlížet (Browse) a případně nějak filtrovat.

A to není zdaleka vše, co musí použitelný MMDBMS zajistit.

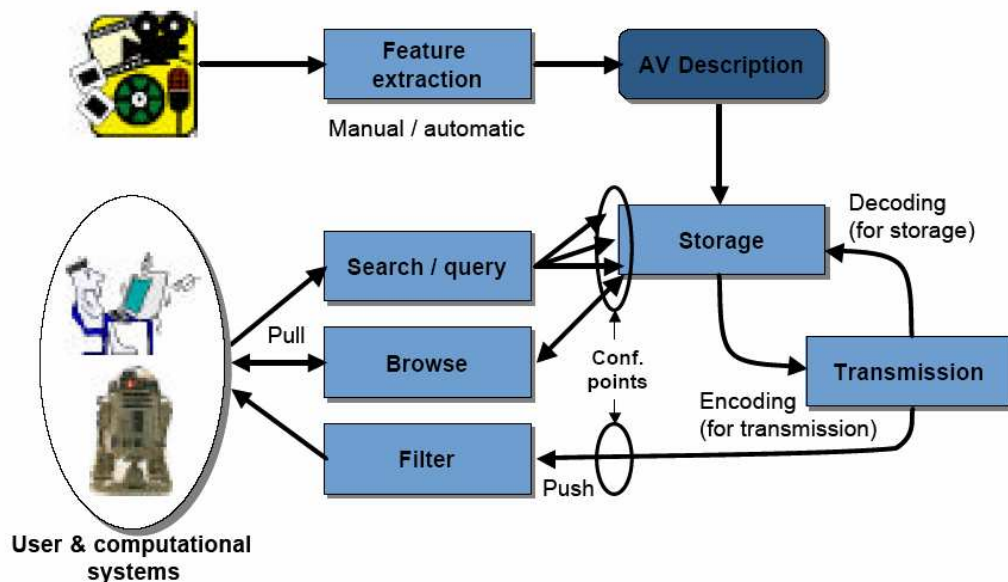


Fig. 2.3 Schéma procesů MMDB (převzato z [MP704])

2.3 Požadavky na konstrukci MMDBMS

Základní požadavky na MMDBMS jsou de facto zděděny z povahy databázových systémů a multimediálních dat. Mezi požadavky na funkčnost DBMS patří modelování,

definice a vytváření slovníku dat, manipulace s nimi, zajištění nezávislosti, bezpečnosti a integrity dat, zotavení po chybách, souběžný přístup, distribuované zpracování a zajištění co nejvyšší výkonnosti [Zen05].

Za vlastnosti zděděné z audiovizuálního povahy MM informací, které mají přímý efekt na konstrukci MMDBMS patří především rozmanitost obsahu, typů a formátů, složitost reprezentace a subjektivní interpretace, rovněž jejich velikost a temporální povaha.

Návrh kvalitního MMDBMS plyne ze splnění následujících požadavků:

- Správa různých typů vstupních, výstupních a úložných zařízení. Vstupem může být mnoho typů dat – WWW, rozmanité dokumenty, obrázky ze skenerů a fotoaparátů, pohyblivé obrázky z kamer. Zvuku z mikrofonů, syntezátorů a MIDI zařízení. Výstupem mohou být jak kvalitní obrazovky s vysokým rozlišením a hi-fi⁵ zvukem, tak jednoduché, mobilní hračky a šumítka. U některých aplikací je ještě více problémů. Například „stejně“ snímky vyhotovené pomocí magnetické rezonance, počítačové tomografie, rentgenu nebo ultrazvuku a k nim náležící textová data.
- Manipulace s množstvím různých datových a kompresních formátů. Jednotlivé MM aplikace obvykle podporují široké spektrum různých formátů. Potřebná komprese způsobuje ztrátu redundantních informací. Problémem je, že se ztratí pokaždé jiná část, respektive množství této informace – barevnost, vyšší frekvence. Například odlišné algoritmy pro změnu rozlišení nebo kompresi obrazu, zvuku a videa.
- Integrace a koexistence datových modelů. Některá data – datum, název, cena a podobné, je výhodnější uchovávat prostřednictvím relačního modelu. Oproti tomu pro dokumenty a MM data je téměř nezbytný objektově orientovaný model. Data je také často nutné uchovávat na externích úložištích.
- Podpora různých platform a operačních systémů pro koncové uživatele. Pro vývojáře aplikací navíc plnohodnotné rozhraní pro několik programovacích jazyků pro tvorbu uživatelských funkcí (extrakce metadat, podobnostní vyhledávání), vývojových prostředí a typů aplikací (podle módy) – binární, webové atp.
- Datové proudy audia a videa mají temporální charakter. Zatímco obraz je nutné prezentovat minimálně 25 (rámců) za sekundu (lépe 30), pro zvuk je to minimálně 8 000 vzorků za sekundu (lépe 22 – 44 kHz) [MP402]. Menší rychlost způsobuje výpadky, „robotické“ zvuky a pohyby, případně zrychlený film jako za Charlieho Chaplina. Proto je nutné splnit daná QoS a real-time omezení.
- Jednotná prezentace výsledků (výsledkem může být libovolný AV obsah nebo jeho popis). Je nezbytná synchronizace z různých typů dat náležejících stejnému MM. Přehrávání pohyblivých obrázků, zvuku a titulků se musí pravidelně synchronizovat, aby se nepředcházeli (vtipné pro prezentace, příjemné pro horory).
- Ochrana autorských práv.
- Nabídka množství jednoduchých dotazů vhodných pro různé typy dat, rychlé a přesné vyhledání požadované informace. Procházení pro začínající uživatele, sofistikované dotazy pro zkušené. Stejně dotazy navíc mohou mít různou podobu – textové (název, žánr), obrazové (foto, náčrtek) nebo zvukové a to specifikací vlastností (melodie) nebo podobným příkladem (ukázka jiné písně).
- Musí být odlišena exaktní povaha parametrů klasických dat (rodné číslo) se subjektivní povahou multimediálních dat (obličej) a to pro každý tip audiovizuálních

⁵ High fidelity – vysoká věrnost zvuku

dat i jejich popis. Liší se navíc podle aplikace. Výsledek podobnostního hledání na dotaz (náčrtek): Klíčové slovo „Zub.“ je pro zubaře vyloženě nanic.

- Vestavěné podobnostní vyhledávání (podle obsahu) musí korespondovat s lidským vnímáním a pojetím podobnosti. To je ale pro každý typ dat odlišné, například barva, tvar, textura, tón, barva hlasu a podobně. Je nutné stanovit jejich vzdálenost.
- Podobnostní vyhledávání ve smyslu časoprostorových dat. Prostorové – topologické a vzdálenostní dotazy (2m vlevo od), temporální dotazy (úsek, kdy zmizel) a jejich kombinace (kde se vyskytoval objekt, který se právě objevil v jiné kameře).
- Je nutné nalézt model dat, tj. vztah mezi syntaxí (popis obsahu, low-level) a sémantikou (význam, high-level) audiovizuálních dat. Také uchování modelů a jejich automatická extrakce z low-level vlastností musí být obsažena v databázi.

Zatímco některé problémy jsou v současné době uspokojivě vyřešeny, jiné jsou stále ve stádiu výzkumu, nebo pouhé vědecké fantazie (sci-fi). Například pro poslední bod platí, že čím speciálnější je aplikace, tím je výsledek uspokojivější. Vyhledání snímků s planetou Saturn, pořízených kupříkladu v srpnu 2006, v databázi hvězdárny, může mít 100% úspěšnost. Oproti tomu nalezení jediného snímku, podobného nalezenému, je na Internetu takřka nemožné.

2.4 Multimedia

Pod multimediálními (MM) daty si lze představit nestrukturovaná data, která dělíme na základě způsobu percepce informace na typ vizuální a audio. Člověk ale nevnímá pouze očima a ušima, ale především mozkiem. Z toho plyne jejich další dělení.

Vizuální data – související se zrakovým vnímáním, jsou následující:

- 2D (planární)
 - Vektorová grafika – data je možné převést na křivky, plochy a podobně (linky, loga, náčrty, mapy), i když mohou být uchována v podobě bitmapy.
 - Text je zvláštní druh vektorové grafiky. Každý znak má svůj význam, je nutné pracovat s kontextem. Někdy bývá zařazen jako vlastní kategorie, ale to porušuje konzistenci s formátovanými dokumenty.
 - Rastrové obrázky (fotky) nemá smysl reprezentovat pomocí vektorů, ale pomocí (3D) objektů a vztahů mezi nimi.
- 3D (prostorové)
 - Modely – geometrické modely reálných objektů (koule, postava).
 - Objekty – reálná instance modelu s jeho vlastnostmi získanými z (2D) rastrů.
 - Ne-objekty – některé oblasti obrázků nejsou objekty (moře, obloha).
- 3-4D (časoprostorové)
 - Pohyblivé obrázky – časový sled obrázků vytváří iluzi pohybu (video bez zvuku).

Audio data se týkají slyšitelnosti:

- Hudba – odlišujeme speciální harmonické vlastnosti zvuků (zvonění).
- Řeč (howk).
- Zvuky nezařazené modelujeme podobně jako obrázky (výbuch).

Pokud pomineme některé speciální případy (genom), pak každý multimediální obsah spadá plně, nebo po částech, do jedné nebo více uvedených kategorií. Například webová stránka obsahuje formátovaný text, (vektorovou) grafiku, rastrové obrázky, ale i videosekvence – pohyblivé obrázky se zvukem (kombinace řeči a hudby).

Jejich kombinace se nazývají audiovizuální (AV) data.

2.5 Formáty audiovizuálního obsahu

Každý digitalizovaný (AV) obsah potřebuje (standardně) definovaný formát pro správnou interpretaci digitálního kódu, který tento kód nemůže(?) nést sebou. Jejich dělení vyplývá ze způsobu (jednoduché) reprezentace pomocí digitální techniky. Často je jeho prezentace nejednotná – různé programy pro každý typ, někdy i formát. Nicméně pro univerzální klienty MMDBMS umožňující prohlížení a tvorbu MM dat, by se měly rozlišovat pouze dva typy – vizuální a audio. Metadata je případně možné reprezentovat jako vizuální data.

2.5.1 Obrázky

Statické 2D obrázky se pro jednoduchost reprezentace dělí na bitmapové (rastrové) a vektorové.

V bitmapě grafice je obraz popsán jednotlivými barevnými, respektive jasovými, hodnotami bodů (pixelů), uspořádaných do mřížky. Jejich zástupci jsou soubory JPEG, PNG, GIF, TIFF, BMP.

Tato reprezentace způsobuje značnou redundanci, proto je výhodné využít nějaký typ komprese. Komprese by měla být bezztrátová, například PNG používá DEFLATE ([RFC 1951](#)). Nicméně lidský mozek citlivý více na jas než na barvu a nižší frekvence než vysoké, proto je možné použít (do jisté míry) ztrátovou kompresi. Jejím zástupcem je především JPEG (Joint Photographic Experts Group), využívající dominantní koeficienty diskretní kosinové transformace (DCT) makrobloků (8x8 px) obrázku v barevném prostoru $YCbCr$ a jejich následnou bezztrátovou RLE (Huffmanovu) kompresi. Další kompresní algoritmy a bitmapové formáty jsou ISO/IEC/ITU-T: JPEG, JPEG 2000, JBIG, nebo PNG, GIF, TIFF, PCX, TGA, BMP, WMP, ILBM, PCX, PSD, ICO.

Vektorová grafika je reprezentována geometrickými útvary, nejčastěji přímkami, křivkami a plochami. Při změně měřítka se neztrácí jejich kvalita. Například SVG, PS, EPS, WMF, AI, CDR, DWG, DXF, HPGL, DVI.

2.5.2 Dokumenty

Dokumenty obsahují více než samotný prostý text (TXT, MIME). Příkladem možných typů dokumentů jsou ODT, RTF, DOC, PDF, TeX, PS, MIF, SGML, HTML, XML. V těchto dokumentech jde kromě kódování textu (ISO/IEC 8859 ASCII, ISO/IEC 10646 Unicode) dále o formátování, strukturu nebo umístění textu, který je doplněn o další multimediální obsah. Existuje velké množství dalších, méně obvyklých dokumentů, například pro publikování (DTP), nebo vývoj (C++, CAD/CAM).

2.5.3 Zvuk

Zvuk je digitalizovaný (1D) audio signál. Je reprezentován diskretními vzorky v časové nebo frekvenční doméně.

Také zvuk obsahuje redundantní informace, bezztrátová komprese ale dosahuje jen 30 – 50% úspory. Nicméně mozek nedokáže, nebo nepotřebuje, všechna data zpracovat. Citlivost lidských uší je (menší než) pásmo 20 Hz – 20kHz. Nosná frekvence hlasu je 50 – 250 Hz, ale v telefonních sítích se používá pouze pásmo harmonických frekvencí 300 – 3400 Hz, které v mozku vyvolávají dojem, že slyší i nosnou frekvenci [Bak87]. Obdobně je to se vzorkovací frekvencí, udává se, že by měla být alespoň 48 kHz (CD

kvalita je 44,1 kHz), ale pro 4 kHz hlasové pásmo je, dle vzorkovacího teorému, dostatečná vzorkovací frekvence 8 000 vzorků za sekundu [Čer05].

Kodeky digitálních signálů pro telekomunikační účely jsou vyvíjeny v rámci ITU (International Telecommunication Union) od roku 1972 G.711 (Siren), následovaly G.721 (ISDN) a G.723, nahrazené G.726 (ADPCM, 1990), G.728 (od 2 kbps, zpoždění 5 vzorků) a G.729 (pro VoIP).

Populární zvukové formáty jsou většinou odvozeny od zvukových stop standardních multimediálních kontejnerů (viz kapitola 2.5.4) i když byly vyvinuty zvlášť (Fraunhofer Institute, <http://www.iis.fraunhofer.de/amm/>).

Příkladem je formát MPEG-1 Audio Layer III (MP3) je komprimován prostřednictvím rozdělení na 32 frekvenčních (MDCT) pásem, ze kterých jsou dle psychoakustických poznatků a dané vrstvy (Layer I - III) vybrány dominantní koeficienty. Kompresní poměr při vzorkovací frekvenci 44100 Hz je 1:5 – 1:11. Současné MP3 používají specifikaci z MPEG-2, resp. MPEG-2.5 (nestandardizováno) pro nízké datové toky.

Vylepšením je zvuk AAC (Advanced Audio Codec) pro datové toky od 2kbps, kompresní poměr pro hlas je tedy až 1:700. Primárně jej využívá až MPEG-4 i když jeho starší verze byla specifikována již v rámci MPEG-2.

Přehled dalších formátů je možné najít například v [Hof03]. Jsou to FLAC, iLBC, RealAudio, WMA, Speex, Vorbis, AIFF, AC3. Popis formátů je v případě volné dostupnosti na stránkách jejich tvůrců (MPEG, FLAC, iLBC). Trendem v komerčních společnostech (RealNetworks, Apple, Dolby Laboratories, Microsoft) je, mimo neustálého vylepšování nosných formátů a software (zařízení) pro jejich prezentaci, poskytovat také multimediální obsah spolu s nástroji pro jejich tvorbu a distribuci.

2.5.4 Video a mediální kontejnery

Video jsou audiovizuální data s temporálním charakterem. To znamená pohyblivé obrázky se zvukem. MM formáty, které nejsou omezeny pouze na tyto data, obsahují navíc alespoň titulky, nebo jiná MM data se někdy nazývají (multi)mediální kontejnery (media container).

Video data obsahují obrovské množství informace, z velké části nadbytečné. Základní idea komprimace je tedy odstranit redundanci v rámci (video)rámců i mezi nimi. Přední kompresní algoritmy byly vyvinuty v rámci spolupráce skupiny odborníků na pohyblivé obrázky MPEG (Moving Picture Experts Group).

Původní formát MPEG-1 (1991) využíval pro odstranění redundance jednotlivých snímků algoritmus JPEG a mezi rámci popisuje především změny, následujícím způsobem. Každých ca. 15 snímků je plný obrázek zvaný I-Frame a mezi nimi jsou popsány případné změny obrazu pomocí pohybu makrobloků (16x16), dopředné prostřednictvím P-Frame a obousměrné B-frame. Datový tok MPEG-1 byl při rozlišení 352 x 288 px (PAL, 352 x 240 px NTSC) asi 1,5 Mbps a dosahoval kompresního poměru 1:50 - 1:100 [Kos03]. Jeho nástupce, MPEG-2 (1996) byl navržen pro digitální televizní a rozhlasové vysílání prostřednictvím satelitů, pozemních vysílačů nebo Internetu.

MPEG-4 (1998) využívá kódování objektů (object coding, jako JPEG2000). Objekty mohou být video, audio nebo 3D (mřížkové) objekty. MPEG-4 umožňuje specifikovat důležitost objektů, například pro zajištění QoS a zavádí mnoho dalších vylepšení kompresních algoritmů pro datový tok od 5 kbps po 1,5 Gbps pro studia, nebo AAC.

Standarty MPEG definují pouze způsob dekomprese, ale v rámci skupiny vyvíjejí i referenční software. Návrh standardů MPEG je blízký digitálním telefonním

standardům ITU, která je formálně přizpůsobuje telekomunikačnímu prostředí. H.261 je příbuzné MPEG1, H.262 MPEG-2, mírně vylepšuje H.263 a H.264 obsahuje stejné části (co se týká kompresního formátu) jako MPEG-4.

Další zástupci formátů videa a MM kontejnerů jsou AVI, AU, ASF, Indeo, Ogg Media, Matryoshka, QuickTime, RealMedia, 3GP, WMV. Mediální kontejnery obsahují mimo vlastního obsahu ještě jeho popis.

2.6 Metadata

Metadata popisují (multimediální) data. Nejen digitálních knihovny vyžadují přesný strukturovaný popis každé uchovávané jednotky [ODL05]:

- Popisná metadata – informace popisující zdroj, jsou viditelná uživateli a slouží především pro vyhledávání. Dle perspektivy uživatelů jsou to [Kos03]:
 - Tvůrce obsahu vloží bibliografické informace, jako je autor, název, shrnutí.
 - Poskytovatel zdroje vloží další metadata, která usnadní nalezení média v jejich kolekci. Například klíčová slova, popis obsahu, index a podobně.
 - Konzument může přidat vlastní poznámky, případně požadavky (na datový tok).
- Administrativní (technická) metadata jsou data nezbytná pro uchování, správu a následné zobrazení obsahu. Obsahují například způsob a čas pořízení, kódování, formát a velikost obsahu, nebo informace o jeho zabezpečení. Jsou záležitostí daného formátu, případně multimediálního kontejneru. Databázový systém by měl umět tyto data interpretovat a využít především pro svou interní potřebu. Pro vyhledávání jsou spíše doplňkové (celovečerní film).
- Strukturální metadata popisují interakci objektů. Například do kterého videa patří daná stopa nebo snímek, stránka dokumentu (kompozice), jejich kategorické uspořádání (agregace), případně jejich provázání odkazy (asociace).

Zbytek této kapitoly se bude věnovat popisným metadatům tvůrce obsahu, následující kapitoly pak i dalším.

Mimo obvyklých informací v hlavičce souboru (typ, rozlišení, kódování, barevnost, vzorkovací frekvence) mohou obsahovat MM data další informace:

- EPG a RDS (Radio Data System) – popis analogového rozhlasového vysílání (např. informace o stanici včetně alternativních frekvencí, skladbě nebo dopravním zpravodajství). Obdobným typem je u digitálního televizního (a rozhlasového) vysílání EPG (Electronic Program(me) Guide), který nabízí více informací o aktuálním i budoucím programu.
- ID3 je popis použitelný výhradně v hlavičce formátu MP3 (MPEG-1 a 2), obsahuje informace jako název, jméno umělce, skladatele, alba, rok vydání a žánr.
- Exif je standard používaný digitálními fotoaparáty při pořizování snímků vyvinutý JEIDA (Japan Electronic Industry Development Association) pro formát TIFF, následně portovaný do JPEG, ale ne už JPEG2000 nebo PNG. Exif obsahuje informace o čase pořízení, informace a nastavení fotoaparátu (značka, typ, citlivost, clonu, expoziční čas, ohniskovou vzdálenost, použití blesku nebo orientaci), informace o místě pořízení (v případě dostupnosti GPS dat), umožňuje uložit náhled fotografie a dodatečně přidání informací o autorovi.
- IPTC (International Press Telecommunications Council) v roce 1979 stanovil několik atributů metadat popisující obrázky, jako název, autor, datum vytvoření, místo pořízení, kategorie, důležitost nebo klíčová slova. Adobe je převzalo, nejprve (1994) vložilo do hlavičky JPEG a TIFF, ale v roce 2005 nahradilo XMP.

- XMP (Extensible Metadata Platform) je open source formát vyvinutý Adobe Systems pro popis obrázků a dokumentů PDF, založený na XML. Zahrnuje informace převzaté z, Exif, IPTC i Dublin Core a doplňuje některé další.
- Dublin core je otevřený standard pro nalezení elektronických zdrojů. Obsahuje základní informace o zdroji (title, creator, subjekt, description, publisher, contributor, date, type, format, identifier, source, language, relation, coverage, rights). Vzhledem k tomu, že byl vytvořen pro anotaci videa, obsahuje rozšíření, které umožňuje sémanticky anotovat jednotlivé scény a podobně. Bývá implementován pomocí XML schématu.
- DICOM (Digital Imaging and Communications in Medicine) vychází ze standardu ACR-NEMA, který popisuje binární formát medicínských souborových dat. Metadata se mohou týkat pacienta (jméno, narozen), jeho stavu, dále zařízení, na kterém byl snímek vytvořen (CT, MR, XR, US⁶, výrobce, nastavení přístroje), poloha pacienta při snímání, a technické informace (barev, rozlišení) a mnohé další. Formát DICOM je kontejner snímků. Umožňuje navíc jednoznačně identifikovat pacienta a uchovává jeho snímky v hierarchii – případ (study, např. rozbitá hlava) obsahuje série (vyšetření), obsahující požadované snímky ARC-NEMA.
- MPEG-7

V přehledu je uvedeno několik běžných formátů pro popis některých typů MM dat. Moderní formáty jako JPEG-2000, PNG, MPEG-4, AAC mají možnost vložit pouze XML metadata (problém s ukazateli Exif), tedy Dublin Core nebo XMP. Ty jsou ale zaměřeny výhradně na (sémantický a technický) popis média ze strany tvůrce a neobsahují žádnou možnost vyhledávání dle obsahu (CBR). Na XML je ale založen i popis dle MPEG-7, který předchází nedostatky doplňuje.

2.6.1 MPEG-7

ISO/IEC 15938: Multimedia Content Description Interface [MP704] je rozhraní pro popis obsahu multimedií. Bylo sice vytvořeno skupinou MPEG, Motion Pictures Expert Group, tedy skupinou expertů na pohyblivé obrázky, ale neslouží pro kompresi audiovizuálních dat.

MPEG-7 poskytuje množinu standardizovaných deskriptorů pro popis obsahu různých druhů médií – statické obrazy v tištěné podobě, 3D grafika a její modely, zvuk, řeč a video nebo biometrie lidského obličeje – s cílem efektivního vyhledávání informací ve velkém množství multimediálních dat.

Popis nesouvisí se způsobem uložení médií. Je možné přiřadit deskriptor MPEG-7 k obrazu, článku v časopise, klasickému filmu, stejně jako proudu MPEG-4.

Pro uložení schémat deskriptorů je definován formát založený na XML, který lze převést do efektivní binární podoby. Norma specifikuje následující prvky:

- Deskriptory (D) reprezentují vlastnosti, rysy – atributy multimediálního obsahu založené na katalozích (název, autor, práva, popis), sémantice (kdo, co, kdy a kde – informace o objektech a událostech), syntaxi (pro CBR, barva obrazu, tón zvuku) a technologii (formát, velikost, vzorkovací frekvence).
- Popisová schémata (Description Schemes – DS) popisují strukturu a sémantiku vztahů mezi komponentami D nebo DS – typ média, jeho původ, možnosti použití, strukturální vlastnosti nebo libovolný text.

⁶ Počítačová tomografie, magnetická rezonance, rentgen, ultrazvuk

- Datové typy (DT). Zavádějí například jednoduchá pole v XML.
- Jazyk pro definici deskriptorů (Description Definition Language – DDL) definuje D, DS, DT, jejich syntax, sémantiku, možnosti jejich změny a rozšíření založené na XML, upravené za pomoci W3C pro MPEG-7. Viz. .
- Systémové nástroje (Systems tools) podporují tvorbu a přenos popisů například v binární podobě, jejich multiplexování s multimediálním obsahem, synchronizaci, formáty souborů.

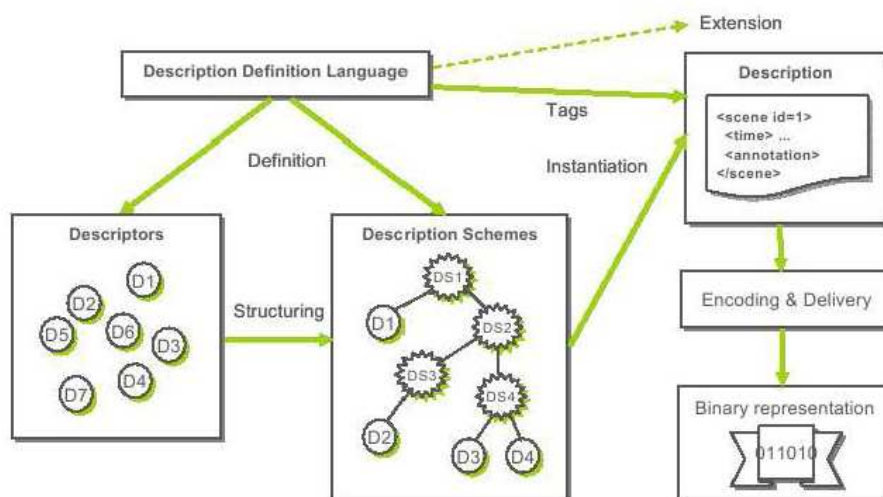


Fig. 2.4 Schéma prvků MPEG-7 (převzato z [7])

Norma MPEG-7 se dělí se na 7 tématických částí:

- Část 1 Systém: specifikuje nástroje pro přípravu deskriptorů k přenosu, ukládání, kompresi a pro synchronizaci s obsahem.
- Část 2 DDL: specifikuje jazyk pro definování standardní množiny deskripčních nástrojů (DS, D, DT) a nových nástrojů, založeno na XML.
- Část 3 Vizualní: obsahuje nástroje pro popis vizuální složky, specifikuje základní kategorie – barva, textura, tvar, pohyb, pozice, rozpoznávání obličeje.
- Část 4 Audio: definuje nástroje pro popis zvukové složky jako spektrální, časové, dynamické vlastnosti (low-level) nebo rozpoznání hlasu, barva nástroje, melodie (high-level).
- Část 5 Schémata popisující multimedia: nástroje pro popis multimediální části popis obsahu – audio i vizuální (video), použití obsahu, organizace, navigace, interaktivity.
- Část 6 Referenční software: implementace standardu se stále vyvíjí, protože norma popisuje způsob uložení popisu, nikoli způsob jeho získání.
- Část 7 Testování shody: poskytuje vodítka pro testování shody implementace deskriptorů.
- Část 8 Extrakce a použití MPEG-7 schémat. Pouze informativní, není součástí normy, například implementace experimentálního modelu.

Detailnější popis MPEG-7 Části 3 a 4 je uveden v kapitole 2.7.

⁷ Pokud by na internetu fungovalo vyhledávání podle obsahu, byla by hračka zjistit, z jakého zdroje pochází tento obrázek...

Příklad popisu obsahu video záznamu dle MPEG-7 je uveden jako Ex. 2.1, namísto manuální anotace DescriptionMetadata (...) je možné umístit Ex. 2.3, jejíž DDL definice je uvedena v příkladě Ex. 2.2

Ex. 2.1 Ilustrace anotace obsahu video dokumentu dle MPEG-7

```
<Mpeg7>
<DescriptionMetadata>...</DescriptionMetadata>
<Description xsi:type="ContentEntity">
<MultimediaContent xsi:type="VideoType">
  <Video id="video_example">
    <MediaInformation>...</MediaInformation>
    <TemporalDecomposition gap="false" overlap="false">
      <VideoSegment id="VS1">
        <MediaTime>

          <MediaTimePoint>T00:00:00</MediaTimePoint>
            <MediaDuration>PT2M</MediaDuration>
          </MediaTime>
          <VisualDescriptor xsi:type="GoFGoPColorType"
            aggregation="average">
            <ScalableColor numOfCoef="8"
              numOfBitplanesDiscarded="0">
              <Coeff>1 2 3 4 5 6 7 8</Coeff>
            </ScalableColor>
          </VisualDescriptor>
        </VideoSegment>
      </TemporalDecomposition>
    </Video>
  </MultimediaContent>
</Description>
</Mpeg7>
```

Ex. 2.2 Ilustrace MPEG-7 DDL pro manuální anotaci video záznamu

```
<complexType name="VideoDoc">
  <element name="Title"/>
  <element name="Producer"/>
  <element name="Date"/>
</complexType>
<complexType name="NewsDoc" base="VideoDoc" derivedBy="extension">
  <element name="Broadcaster" maxOccurs="1"/>
</complexType>
<element name="VideoCatalogue">
  <complexType>
    <element name="CatalogueEntry" type="VideoDoc"/>
  </complexType>
</element>
```

Ex. 2.3 Ilustrace manuální anotace video dokumentu pomocí MPEG-7

```
<CatalogueEntry xsi:type="NewsDoc">
  <Title>"CNN 6 oclock News" </Title>
  <Producer>David James</Producer>
  <Date>1999</Date>
  <Broadcaster>CNN</Broadcaster>
</CatalogueEntry>
```

2.6.2 MPEG-21

MPEG-21 je nastupující standard pro komunikaci mezi všemi uživateli – tvůrci, distributory a konzumenty multimediálního obsahu v distribuovaném multimediálním systému [MP210]. Jednotkou dat je unikátní digitální položka (Item), obsahující AV obsah (JPEG2000, MPEG-4 a další) spolu s popisem MPEG-7, správou autorských práv a systémem pro jejich distribuci, prezentaci a synchronizaci.

Standard byl v době tvorby této práce (2006) za polovinou schvalovacího procesu.

2.7 Extrakce rysů a popis obsahu

V předchozí kapitole bylo uvedeno, jak vypadá popis multimediálních dat. Tato je věnována způsobu, jak jej získat. V seznamovací kapitole (2.2 Multimediální databáze) je rozlišen manuální a automatický popis obsahu multimédií. Manuální popis není třeba příliš rozebírat. Jde o vytvoření smysluplného, objektivního názvu a popisu objektů, hovoru, scén a podobně, případně o ruční přiřazení média do jisté kategorie.

Automatický popis vytváří vektor rysů⁸ (deskriptor, signaturu) pomocí procesu známého jako extrakce rysů⁹. Rysy jsou jisté charakteristické vlastnosti obsahu média. Rysy člověka mohou být například barva kůže, vlasů, očí, jejich vzdálenost, velikost nosu, hlavy, drsnost kůže atd. Podobné jsou i rysy obrázků. Rysy pohyblivých obrázků mohou být pohyb člověka nebo jeho jiné chování. Rysy textu jsou až na výjimky (rozpoznání jazyka, autora) klíčová slova. Zvukové rysy jsou například frekvenční vlastnosti, tón, barva nástroje, hlas, rozpoznání řeči a podobné informace obsažené v MM datech.

2.7.1 Popis obsahu audia

Na rysech audia je nejlépe možné rozlišit jejich úroveň:

- **Nízká úroveň popisu** (low-level) zahrnuje množinu rysů společných pro všechny signály. Extrakce spektrálních, temporálních a dynamických (syntaktických) vlastností zvuku je plně algoritimizovatelná. Nejčastěji se používá rychlá Fourierova transformace (FFT) [Čer05] pro spektrum zvuku (Fig. 2.5), určení harmonického spektra, hustotu výkonu, základního tónu a jeho zabarvení¹⁰.
- Střední úroveň (mid-level) se někdy označují statistické modely rysů dat. Běžně se počítají statistiky (střední hodnota, korelace) nízkourovňových rysů, které mohou sloužit pro podobnostní vyhledávání, dotazem je například krátká ukázka. Je možné určit, zda se jedná o pravidelnou melodii, řeč, při absenci základního tónu je možné detekovat šum nebo ticho. Dále je možné odlišit zvukové segmenty (zprávy, hudba).

⁸ Feature vector (descriptor, signature)

⁹ Feature extraction

¹⁰ Timbre, zabarvení hlasu nebo zvuku hudebního nástroje pomocí harmonických frekvencí

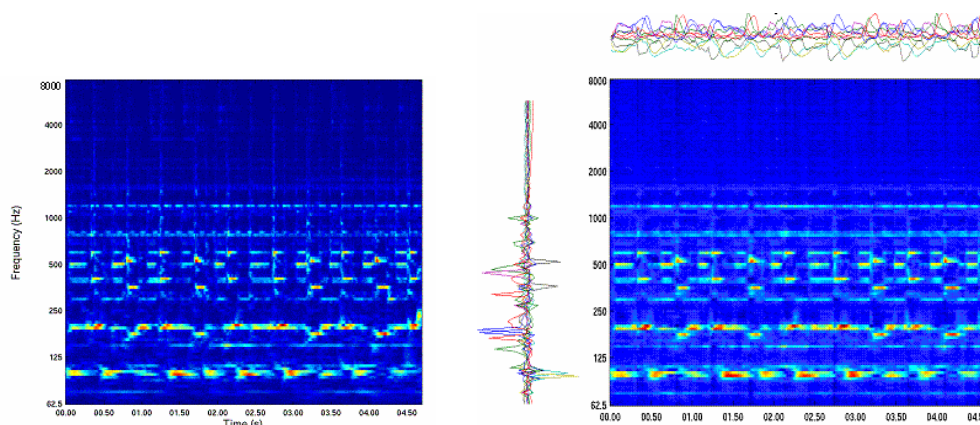


Fig. 2.5 Ukázka spektrálních (low-level) vlastností popu a rocku [MP704]

- Vysoká úroveň** (high-level, deep-level) je obvykle získána z rysů nižší úrovně. Jedná se například o sémantickou anotaci několika vzorků v databázi, ze kterých jsou extrahovány rysy, analyzovány nějakým učícím se algoritmem (Gaussovské, Bayessovské, Markovovy modely) a následně slouží pro klasifikaci (zvuku), jako v kapitole 2.15. Tyto procesy jsou poměrně dobře stanoveny pro speciální aplikace, jako je rozpoznání hudebního nástroje, melodie a především pro rozpoznání mluvčího a obsahu řeči. Dotazem tedy může být zantování melodie (query-by-humming), nebo požadovaná slova, jako by se jednalo o text.

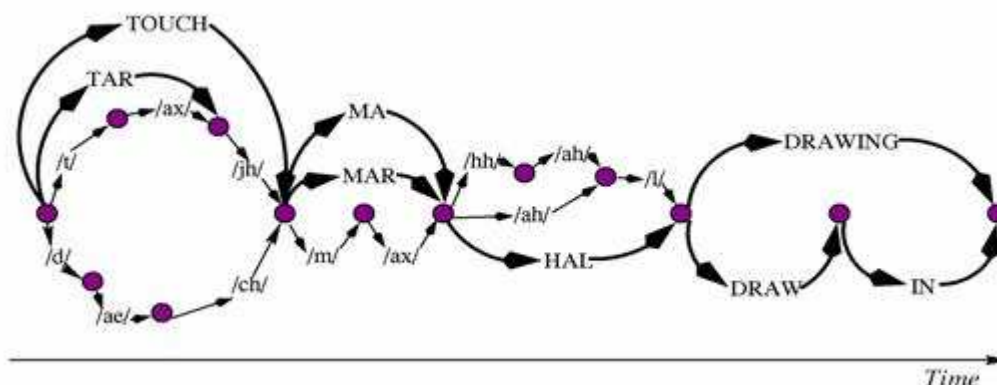


Fig. 2.6 Proces rozpoznání řeči (high-level) „Taj Mahal drawing“ [MP704]

Systémy pro automatické rozpoznání řeči (Automatic Speech Recognition, ASR) jsou schopny identifikovat „naučená“ slova pomocí jejich databáze (slovníku) a odpovídajícího repertoáru fonémů (hlásky, slabiky). V příkladě na obrázku Fig. 2.6 je ilustrován proces rozpoznání řeči s jedním slovem neznámým slovem a druhým slovníkovým. Jsou nalezeny uzly (mezi fonémy) a k nim je vytvořen (alternativní) graf fonémů a odpovídajících slov, ze kterého je vybrán nejpravděpodobnější přepis audio obsahu dle naučených, anotovaných řečových dat.

2.7.2 Extrakce rysů z obrázků

Nízkoúrovňové rysy obrazových dat jsou relativně jednoduché. Jedná se například o dominantní barvu v obrázku nebo její histogram. Pro tyto rysy není barevný model RGB ani CMYK příliš vhodný. Je rozumné použít model, který lépe odpovídá lidskému chápání barev, například HSV, HLS, nebo alespoň $YCbCr$ [Krš05].

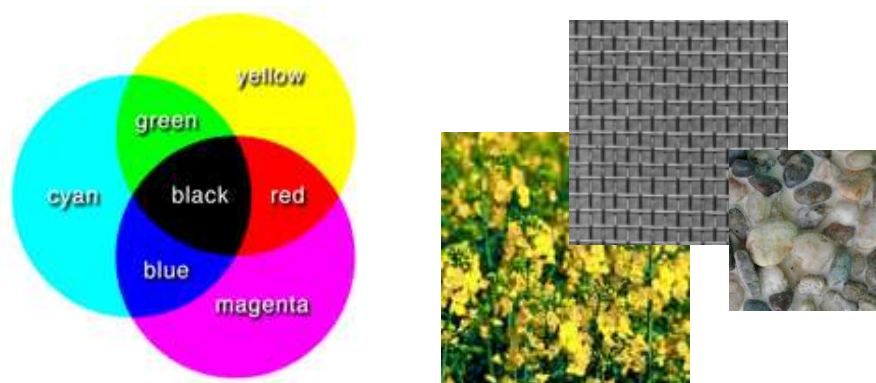


Fig. 2.7 Ilustrace barev a textur

Textura je vlastností všech reálných povrchů. Má 2D vizuální vlastnosti, například kontrast, hranová frekvence, zaměřenost, pravidelnost, nebo hmatový charakter, jako hrubost, drsnost nebo huňatost. V její počítačové reprezentaci je složená z opakujících se elementů, které se nazývají primitiva. Tato primitiva mohou být popsána jejich statistickou, geometrickou nebo frekvenční charakteristikou.

Jedním z nejefektivnějších způsobů popisu texturní informace je extrakce rysů z frekvenční oblasti, získané například DCT nebo FFT, protože odpovídá lidskému chápání. Velikost energie, rozdělené do (30) frekvenčních kanálů – dle jejich směru (6, po 30°) a velikosti (5, po oktávách), slouží jako vektor rysů (30 hodnot). Tato technika se nazývá Gaborovy filtry [Chm04], které jsou zobrazeny na obrázku Fig. 2.8.

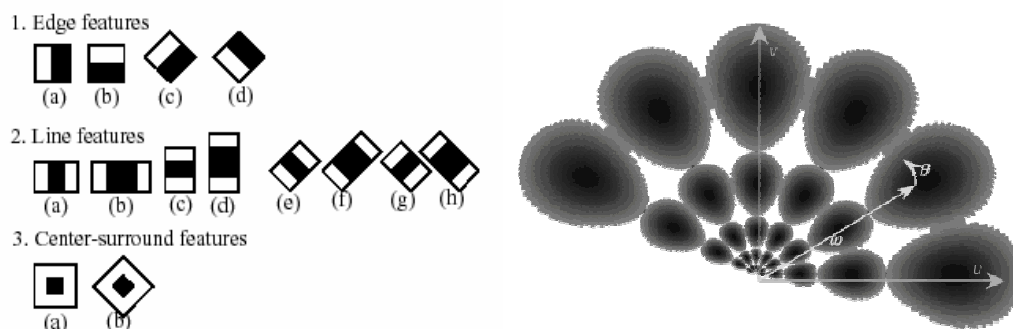


Fig. 2.8 Ukázka Haarových waveletů a Gaborových frekvenčních filtrů

Wavelety, nebo podobné rysy, jsou také běžně používané pro tvorbu deskriptorů. Mají prostorově-frekvenční charakter. Ukázka filtrů, se kterými může být obraz konvoluován je zobrazena na obrázku Fig. 2.8. Jejich výsledek je možné statisticky zpracovat a při vhodné kombinaci tvarů – rysů (například Haarovy kaskády), je možné nalézt také požadovaný objekt v obrázku.

Tyto vlastnosti jsou samy o sobě vhodné například k optické kontrole výroby pro laky automobilů nebo nanotextilie. Barva a textura jsou vhodné pro podobnostní vyhledávání objektů v databázi střední úrovně, i když nedávají vždy korektní výsledky, pokud se vztahují k celému obrázku, namísto jednotlivých objektů. Libovolný obrázek představuje mozaiku barev, lépe textur. Proto je velmi efektivní indexovat oblasti s danou texturou nebo barvou, i když by se mělo jednat pouze o čtvercové oblasti, pevné velikosti, shodné pro všechny obrázky.

2.7.3 Tvar

Hlavním problémem vizuálních dat je, že jsou příliš rozmanitá. Objekty, tak jak je vnímá člověk, jsou odlišné od způsobu jejich rozpoznání počítači, tedy oblasti s jistou barvou,

texturou, hranami a podobně. Mohou mít různou velikost, orientaci nebo barvu a přitom mají stejný význam, nebo naopak, například auto.

Existuje ale několik světlých výjimek, například analýza snímků vzdáleného pozorování země takto umožní relativně spolehlivě klasifikovat osídlené oblasti, pole i lesy a určit, zda nejsou napadeny například kůrovcem a vymežit toto napadení.

Pokud se oprostíme od způsobu, jak přesný tvar objektu získat (viz. Pohyb 2.7.4, nebo Získávání znalostí 2.15), je možné říci, že objekt se skládá z jednoho nebo více regionů v obrázku. K tomu počítáme i díry či další přerušování oblastí. Oblasti je možné popsat jako plochu, nebo její hranici:

- V případě plochy bývá oblast reprezentována jako nějaká malá bitmapa (u MPEG-7 17.5 bytů), ve které hodnoty (jasu) 1 značí příslušnost do odpovídajících oblastí, zatímco 0 k bílému pozadí. Tento způsob popisu oblasti je efektivní pro porovnávání, sledování ve videu a v případě nějakého porušení hranice.
- Hraniční reprezentace je náročnější, ale lépe odpovídá lidskému vnímání a je více odolná vůči afinním transformacím [Krš05], nehledě na možnost změny tvaru objektu ve videu, například běžícího člověka [MP704].

V současné době sice není extrakce 3D tvaru objektů ze 2D příliš rozvinutá, nicméně se hodí pro uchování modelů – předpisů, jak má reálný objekt ideálně vypadat. Jejich reprezentací je obvykle 3D mřížka.

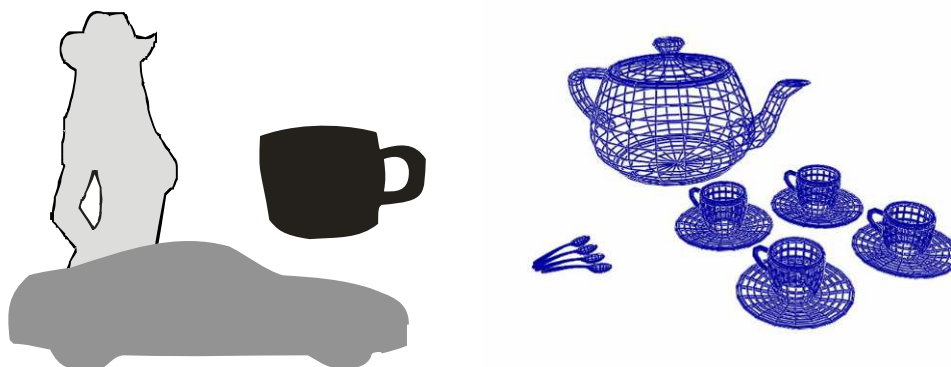


Fig. 2.9 Reprezentace tvaru plochou, hraniční křivkou a 3D mřížkou

2.7.4 Pohyb a umístění

Pohybem je myšlena změna umístění objektů v rámci následujících snímků i další transformace objektu, jako přiblížení, zmenšení, otočení, perspektivní deformace. Dalším typem pohybu, je pohyb kamery.

Kamera má obecně šest stupňů volnosti – může se pohybovat po svislé i vodorovné ose, přibližovat se a vzdalovat (zoom) a v každé z těchto os může rotovat. Pohyb kamery není vždy jednoduché zjistit (například v davu lidí), ale je to žádaná vlastnost. Pohyb kamery lze vyjádřit pomocí fundamentální matice. Tu lze odhadnout pomocí shody několika náhodných bodů v následujících snímcích, například pomocí RANSAC (Random Sample Consensus) [Chl05].

Obecným typem pohybu objektů na ploše je perspektivní transformace, která se běžně využívá například při fotbalových přenosech. Její výpočet je jen o trochu jednodušší, než výpočet fundamentální matice [Jäh99]. Argumentem složitějšího počítání, je možnost zjistit reálnou polohu objektů ve scéně, která je nezbytná například pro bezpečnostní aplikace založené na kamerových systémech.

Pokud si nepotřebujeme komplikovat život, nejjednodušším typem pohybu tuhého objektu je pohyb v rovině, kdy připouštíme pouze translaci (jako MPEG-2) a někdy i rotaci objektu. Těleso tedy příliš nemění vzhled, tvar ani velikost. Takováto tělesa lze ve videu najít pomocí metod pro estimaci (odhad) pohybu. Jednodušší metody zjistí změnu v po sobě následujících snímcích tím, že odečtou hodnoty jejich pixelů. Pak pracují buď po blocích (16 x 16 px), nebo ohraničí celou oblast a snaží se v blízkém okolí najít její nový výskyt. Tím vytvoří jednoduchý vektor pohybu. Lépe funguje například prediktivně – korekční Kalmanův filtr [Jäh99].

Když kompenzujeme pohyb kamery (odečteme vektor pohybu) a ohraničíme dále se pohybující oblasti (seskupíme bloky), jsme schopni identifikovat pohyblivé objekty a určit jejich tvar. To je velmi ceněná vlastnost při anotace pohyblivých obrázků.

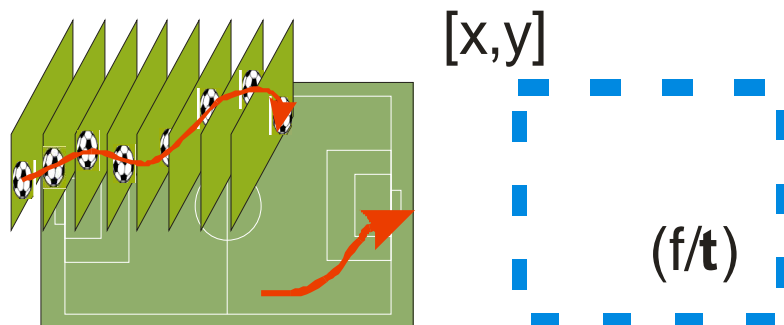


Fig. 2.10 Pohyb a umístění

Takovéto vizuální těleso je poté možné obalit konvexní obálkou (obdélníkem) a zaznamenat jeho souřadnice do databáze. Pokud doplníme časové razítko (číslo rámce), získáme časoprostorový lokátor (umístění) objektu. V případě nalezení nového objektu můžeme udělat více, než zaznamenat jeho velikost. Od manuální anotace (jméno herečky), přes extrakci rysů nižších úrovní až po automatické rozpoznání (viz. kapitola 2.15). A pokud máme k dispozici i informace o reálném prostředí, lze umístění objektu anotovat v jeho kontextu a uchovávat v (čas)prostorové databázi.

2.7.5 Speciální aplikace

Různé speciální aplikace multimediálních databází vyžadují odlišné popisy dat, ať už se jedná o jejich popis nebo o popis obsahu. Těžko může nějaká (tato) práce nebo norma obsáhnout všechny doménově specifické rysy. Z toho důvodu je doporučeno uchovávat metadata ve formátu XML (eXtensible markup language, rozšiřitelný značkovací jazyk).

Zajímavou aplikační oblastí je také nalezení a rozpoznání obličeje. Detekci je hlavy (demo) je možné udělat například dle barvy pokožky, ale ukazuje se, že texturní rysy obličeje, jako waveletová transformace nebo Gaborovy filtry dávají lepší výsledky [Hin2003]. Po ohraničení obličeje (obvykle se předpokládá pohled zepředu) jsou nalezeny dominantní části jako oči a ústa. Z těchto dat je možné určit některé biometrické informace jako poměry vzdálenosti očí, úst, nebo velikost nosu či čela. Běžnější je ale extrakce texturních rysů, nebo umístění obličeje do mřížky s pevnou pozicí očí a popis obličeje pomocí vektoru složeného z průměrné barvy (intenzity) každé buňky [MPEG-7].

Jako další příklad uvedu bezpečnostní aplikaci, která má za úkol měřit průměrnou rychlost jízdy pomocí dvou kamer, na začátku a konci měřeného úseku. V tomto případě není třeba složitě popisovat pohyb a umístění automobilu, protože ze 2 uložených fotografií s časovým razítkem je vše zřejmé. Automobily, které nepřekročí rychlost (o hodně), nás nezajímají a nejsou zaznamenány. Nejvýše jejich počet.

Obdobné je to v případě dalších aplikací (ukázka na obrázku Fig. 2.11), kde je možné obecné informace analyzovat a ihned zahodit, protože pro danou aplikaci nejsou třeba a běžné podobnostní vyhledávání by stejně nedávalo relevantní výsledky. Budeme především chtít vyhledávat podle značky vozidla, které rychlost překročilo.

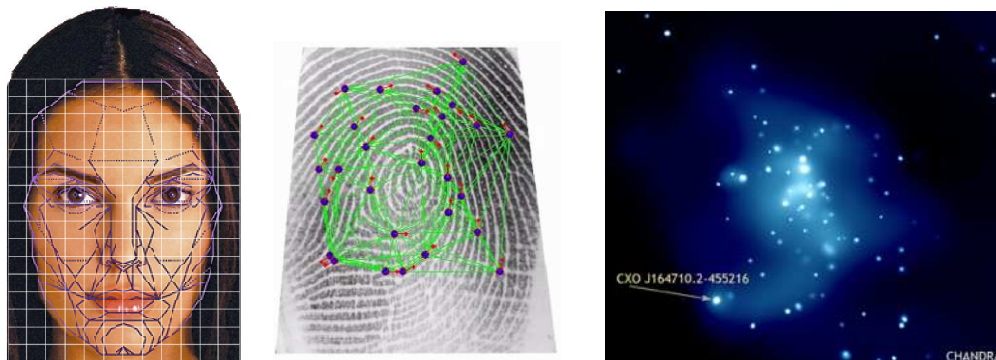


Fig. 2.11 Speciální aplikace – rozpoznání obličeje, otisku prstů, nalezení (neuronových) hvězd

2.8 Vyhledávání

Vyhledání relevantní informace v audiovizuálních datech je značně náročný, ale elementární úkol každého MMDBMS. Z toho důvodu požadujeme, aby bylo vyhledávání co možná nejefektivnější. Proto je v naprosté většině případů omezeno na metadata (obsahující popis i rysy), která musí být předem extrahována a vhodně indexována, na čemž výrazně závisí efektivita vyhledávání (Fig. 2.13). Vyhledávací algoritmus je také důležitý, ale především musí umět interpretovat správně daná metadata. Například hodnoty 0 a 1 představují v masce tvaru různé objekty, kdežto v případě jasu jsou to de facto stejné hodnoty (z 256).

Metadata jsou relativně běžná, strukturovaná (nejhůře XML) data. Je tedy možné použít libovolný strukturovaný dotaz. Tak se tomu děje v případě vyhledávání dle textového popisu dat, které je v současné době asi nejběžnější (Obrázky na Google). V příkladě Ex. 2.4 se oprostíme od prezentace obsahu.

Ex. 2.4 Dotaz dle popisu dat

```
select i.content  
from Images i, Description d  
where i.id = d.image and  
d.annotation LIKE '%Mars%';
```

Druhým způsobem je vyhledávání dle obsahu audiovizuálních dat. Pro něj je výše popsaný způsob krajně nepraktický. Například při dotazu na nalezení jisté textury by bylo nutné znát všech 30 koeficientů deskriptoru. Takovýto způsob se nazývá vyhledávání specifikací vlastností.

Některé MMDBMS dokáží specifikaci vlastností výrazně zjednodušit. Je možné například dotaz načrtnout, tj. specifikovat barvu a tvar objektu (query by sketch), nebo zanotovat melodii (query by humming). Příkladem je systém QBIC (<http://www.qbic.almaden.ibm.com/>)¹¹. Přestože je tento způsob pro lidi velmi intuitivní, netěší se takové podpoře, jako podobnostní vyhledávání, se kterým má mnoho společného (ilustrace Fig. 2.2).

¹¹ <http://www.hermitagemuseum.org/cgi-bin/db2www/qbicSearch.mac/qbic?selLang=English>

2.8.1 Podobnostní vyhledávání

Člověk umí vyjádřit podobnost jistým „nativním“ způsobem. Ta může být velmi subjektivní (myš), ale i více objektivní (zebra). Podobnost proto musí brát v úvahu především modely lidského vnímání [Hig06].

Princip podobnostního vyhledávání je relativně jednoduchý. Předpokládá, že pro vyhledávání požadované informace jsme schopni dodat nějaký příklad požadovaného audiovizuálního obsahu. Z tohoto média MMDBMS extrahuje stejné rysy, jako by jej chtěl uložit do databáze. Tato metadata dotazu jsou poté využita pro vyhledávání (viz. schéma Fig. 2.3).

Běžně se používají první dvě podobnostní funkce:

- Pomocí tříd podobnosti. Máme-li množinu podobných objektů, pak relace podobnosti na této množině je reflexivní (soběpodobnost) a symetrická (jako dvojčata). Ale objekt této množiny není podobný objektu jiné množiny. Formálně viz [Češ06]. Objekty mohou být označeny (tag, label) například ručně. Přiřazení do těchto množin, případně způsob jejich automatické tvorby je naznačen v kapitole 2.15.
- Pomocí (Eukleidovské) vzdálenosti (rysů). Máme-li spojitě (číselné) intervaly, a je-li možné mezi nimi měřit vzdálenosti (měla by platit symetrie, nezápornost a trojúhelníková nerovnost), pak můžeme tuto vzdálenost označit jako míru podobnosti. Nicméně záleží pouze na návrhu vektoru rysů, jestli bude tato podobnost odpovídat také jejímu lidskému pojetí. Např. frekvenční spektrum ($\alpha_i = 1$) versus délka záznamu ($\alpha_j = 0$). Pro vektory v_1, v_2 o N rysech lze vyjádřit podobnost jako váženou (α) vzdálenost d :

$$\text{Eq. 2.1} \quad d(v_1, v_2) = \sqrt{\sum_{i=1}^N \alpha_i (v_2[i] - v_1[i])^2}$$

- Podobnost odvozená od (čas)prostorových dat [Kol05]. Prostorové pro topologické a vzdálenostní dotazy (2m vlevo od). Temporální dotazy (úsek, kdy zmizel). A jejich kombinace (sledování více kamerami, kde se vyskytoval objekt dříve).
- Transformační vzdálenost¹².

Podobnostní vyhledávání v (čas)prostorových datech má několik úskalí. Tím prvním je jejich spojitý charakter – objekty je nutné vyhledávat pomocí plovoucího okna. Druhým je, že objekt může být různě velký a znamenat to stejné (portrét, dav) nebo naopak. Proto se používají algoritmy s adaptivní velikostí okna, které jsou invariantní velikostně (adaptive scaling) [Gra04] a časově (dynamic time warping) [Kad02].

Příkladem vyhledávání v obsahu audio dat obsahujících muziku je http://www.music-ir.org/mirex2006/index.php/Main_Page. V případě řečových dat je to například skupina Speech@FIT (<http://www.fit.vutbr.cz/research/groups/speech/>), nebo software <http://www-306.ibm.com/software/voice/viavoice/>. Pro obsah videa jsou to například produkty: <http://www.almaden.ibm.com/projects/cuevideo.shtml>, <http://www.virage.com/>, <http://persia.ee.columbia.edu:8080/>.

- ¹² Je popsána v literatuře. Podobnost dvou obrázků je vyjádřena cenou transformace prvního na druhý nebo naopak [Sub98]. Transformace pracují přímo s obsahem pro porovnání každé dvojice obrázků, což je silně neefektivní. Výhodou je možnost vyjádřit subjektivitu. Jedná se pouze o akademickou práci, realizace mi není známa.



Fig. 2.12 Ukázka podobnostního vyhledávání (<http://corbis.ltutech.com/>)

Příklad podobnostního vyhledávání obrazových dat (jako Fig. 2.12) je možné nalézt v následujících lokacích: <http://www.tiltomo.com/>, <http://www.pyxor.com/>. Software pro podobnostní vyhledávání je například: <http://sourceforge.net/projects/brisc>, <http://octagon.viitala.eu/>. Demonstrace podobnostního vyhledávání pomocí MMDB rozšíření Oracle Database je uvedena jako Příloha 2. Oracle interMedia aplikace.

2.9 Indexace

Index je prvek databáze, který má na starost rychlý přístup k datům a jejich vyhledání. Indexovat lze popis i obsah. Na obrázku Fig. 2.13 je znázorněna motivace využití indexu při dotazování obsahu obrázkových dat.

Použití indexu (pokud existuje) určuje plánovač na základě optimalizovaného dotazu. Optimalizátor a plánovač je v databázových systémech odvozen od standardních. Vzhledem k velkému množství typů MM dat a metadat, jejich indexů, dotazovacích jazyků (relační, objektově-relační, objektové) a množstvím různých produktů, nicméně komplikovaný. Základem jejich MM přizpůsobení je především nezpracovávat data, která (v daný moment) není třeba, nebo nelze prezentovat.

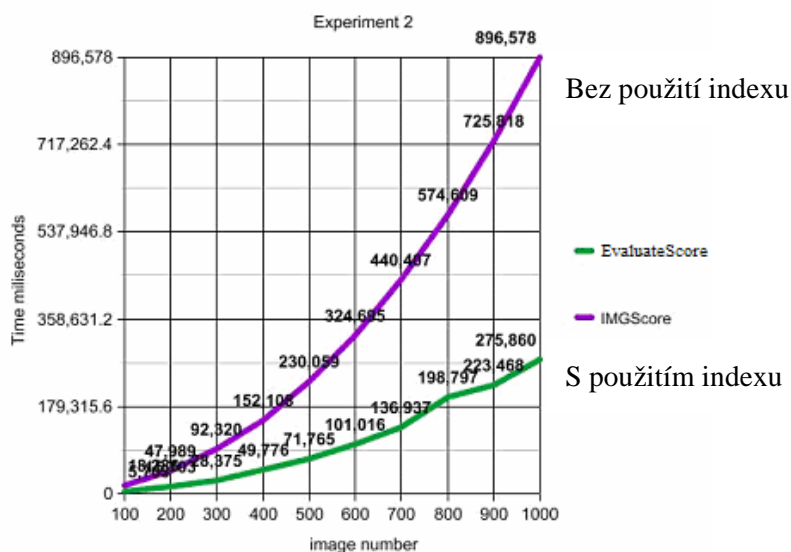


Fig. 2.13 Motivace využití indexu v databázi Oracle interMedia [Gir06]

Nejjednodušším způsobem indexování obsahu je užití předepsaných klíčových slov (label), například hokej, rock. V běžném textu se slova mohou opakovat a je nutné uvažovat jak synonyma, tak homonyma (řeší se stromem).

Nejznámější strukturou pro indexování textových dat, je frekvenční tabulka. Řádky představují jednotlivá klíčová slova (termy), což jsou všechna slova vyskytující se v uložených dokumentech (wordlist), často bez nejběžnějších slov (a, to) zvaných stoplist, nebo je použit jen kmen slova (slov). Sloupce pak tvoří jednotlivé dokumenty, respektive počet klíčových slov v daném dokumentu. Je možné použít její modifikaci pomocí řídké tabulky, jak je naznačeno v tabulce Tab. 2.1.

Tab. 2.1 Ilustrace řídké frekvenční tabulky

(Kmen) slova	Dokumenty obsahující slovo (počet)
Myš	Veterina.odt (39), Deratizace.pdf (19), Periferie.xls (3)

Indexovat lze i další obsah, například prostřednictvím vektoru rysů. Pokud jsme schopni vytvořit vektor rysů s konstantním počtem číselných parametrů (barevný histogram, textura), je možné použít klasické indexové struktury, jako jsou B^+ stromy nebo bitmapový index [Zen05].

Tradičním požadavkem MMDBs je indexování (anotovaných) objektů. Problémem multimediálních dat je automatické nalezení těchto objektů (nastíněno v kapitole 2.15) a především jejich vícerozměrná povaha. Nelze tedy použít klasické metody, ale je možné použít struktury založené na dělení prostoru.

2.9.1 K-D stromy

K-D stromy slouží pro ukládání bodových k-dimenzionálních dat. Dělí prostor podle vkládaných bodů. Při stavbě stromu indexu vložíme první bod. Do levé části stromu dáme všechny další body, které mají menší hodnotu souřadnice x a do pravé větší. Při vkládání dalšího bodu se prostor rozdělí podle další souřadnice viz Fig. 2.14. Nevýhoda K-D stromů je špatné pořadí při vkládání, výhodou je snadná implementace.

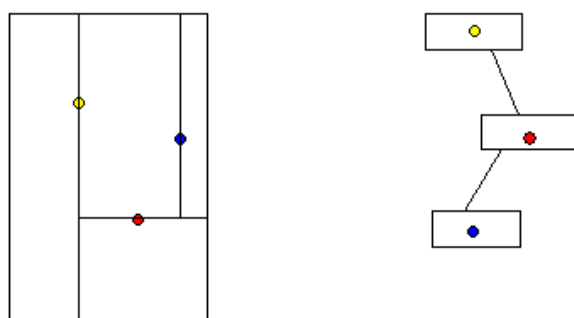


Fig. 2.14 Dělení 2D prostoru pomocí K-D stromů [Nos06].

2.9.2 Quad stromy

Quad stromy vycházejí z K-D stromů, ale vkládané body rozdělují prostor ve všech směrech, viz Fig. 2.15. Quad stromy umožňují oproti K-D rychlejší vyhledávání, ale operace rušení je náročnější, kvůli nalezení uzlu, který má být nahrazen.

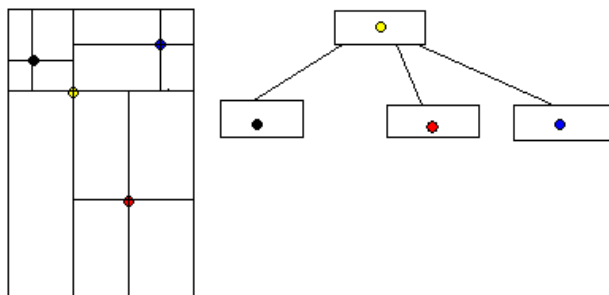


Fig. 2.15 Dělení 2D prostoru pomocí Quad stromů [Nos06].

Jejich modifikací jsou MX-quad stromy. Také dělí prostor podle všech souřadnic, ale jen tak, aby vzniklé části byly stejně velké. Dále dělí rekurzivně. V těchto stromech je pevně určena granularita, problémem je její efektivní určení.

Zobecnění K-D a Quad stromů jsou BSP (Bojary Space Partition), které obdobně rozdělují prostor libovolnými rovinami na konvexní oblasti, obsahující maximálně jeden objekt, případně jeho část.

2.9.3 R-stromy

R-stromy jsou založeny na ukládání obdélníkových struktur. Některé obdélníky představují reálné objekty vložené do databáze. Další sdružují menší obdélníky do větších, viz Fig. 2.16. Všechny uzly stromu, mimo kořenového, musí být alespoň z poloviny zaplněné. Pro neobdélníkové objekty se hledá minimální ohraničující obdélník (Minimum Bounding Rectangle, MBR). Modifikací R-stromů jsou M-stromy s kruhovými (kulovými) oblastmi, což je výhodnější pro určování eukleidovských vzdáleností, ale výpočetně náročnější.

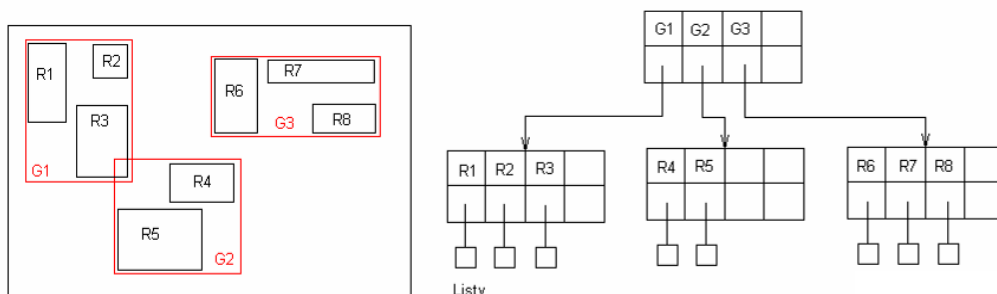


Fig. 2.16 R-stromy a dělení 2D prostoru z [Zen05].

R-stromy jsou nejvhodnějším algoritmem pro velké objemy dat. Problémem je, že se obdélníky mohou překrývat. Při vyhledávání potom nelze vyloučit určité větve, ale musí se prohledávat větší část stromu.

To řeší lépe až jejich modifikace. R^+ stromy vylučují překrývání, ale obdélník reálného objektu může být zařazen do několika listů. Běžnější jsou R^* stromy, které se při výběru uzlu pro zařazení a štěpení snaží překrytí minimalizovat [Sub98].

2.10 Ukládání

Základní vlastnost databáze je uchování dat. Pro obrovské objemy multimediálních dat to představuje značné problémy. Zatímco pro klasická strukturovaná data je jejich umístění v datovém souboru s transakční kontrolou samozřejmé, pro multimedia je kvůli

zvýšeným výkonnostních, fyzickým a finančním důvodům situace složitější. Mohou být uloženy uvnitř databáze, na streamovacím serveru anebo jinak distribuovaně:

- Uvnitř databáze jako objekt (OO DB), nebo jako BLOB (Binary large Object) v relační databázi s transakční kontrolou.
- V externím binárním souboru (BFILE), na diskovém poli nebo streamovacím serveru, který spravuje pouze DBMS, s transakční kontrolou.
- V externím souboru (souborech), ke kterým nemá MMDBMS výlučný přístup, je možné transakční kontrole podrobit pouze ukazatele (jména souborů).
- Na libovolném HTTP, FTP nebo RTPS serveru. Uchovává se jen URL.
- Na jiném (distribuovaném, streamovacím) zařízení, např. kamera s dočasnou pamětí.
- Na „pružném“ či optickém disku, pásce, nebo jiné paměti, bez možnosti okamžitého vyhledání a doručení.

2.11 Přenos

Z hlediska lidského vnímání je 1/4s ztráta (zamrznutí) video signálu omluvitelná, zvukového už ale těžko. V reproduktorech nepříjemně zapraská nebo se začne cyklicky opakovat poslední přijatá frekvence. Fakt, že MM SRBD musí zajistit přenos multimedií pro plynulé přehrávání s dostatečnou kvalitou je dnes samozřejmostí.

V tomto ohledu je transport dokumentů a statických obrázků jednoduchou záležitostí (prostřednictvím HTTP, FTP). Nicméně i jejich přenos je vhodné optimalizovat. Příkladem je progresivní kódování JPEG, kde jsou nejprve zasílány hrubé informace o vzhledu celého obrázku a postupně upřesňovány.

Na distribuci dalších (real-time, continuous) multimediálních dat bývají kladeny různé požadavky, proto je jejich obsah distribuován více způsoby. Obecně je možné je zařadit dle míry jejich interakce:

- Žádná interakce. Do této kategorie spadá televizní a rozhlasové vysílání v libovolné formě – pozemní, satelitní vysílání, obecní rozhlas nebo internetové stanice. Jediná možnost interakce je stanici (program, kanál, stream) vypnout, resp. přepnout. Proto se používá plošné vysílání (broadcast), v Internetu multicast.
- Obsah přístupný na přání (na vyžádání, on demand). Řeší MMDBMS např. vyrovnávací paměť. Pokud nepožadujeme vyhledávání, lze použít streaming server.
- Interaktivní komunikace. Příkladem jsou telefonní hovory nebo videokonference. Jsou vyžadovány speciální protokoly, komunikační aplikace a (DB) servery pro navázání spojení, které může být přepínané, ale i paketové.

S přenosem multimedií souvisí nutnost v síťovém prostředí zabezpečit dostupnost dat využitím vyrovnávacích pamětí a dostatečné přenosové kapacity. Především z důvodu real-time přenosu multimediálních dat spolu s klasickými byla vyvinuta přepínaná síť ATM. Vzhledem k vyšší ceně ale nebylo její nasazení široce realizováno. Pro úsporu přenosového pásma byly vyvinuty také protokoly nad IP, zejména pro multicast, viz předměty PDS [Švé05] a PSI. V současné době je jejich hlavním problémem absence univerzálního mechanismu pro zajištění dostatečné přenosové rychlosti (Quality of Service, QoS).

Audiovizuální obsah je nutné často přenášet prostřednictvím nejrůznějších technologií s různou přenosovou kapacitou (DVB, LAN, DSL, 3GP nebo GPRS¹³). Je tedy nutné

¹³ Digitální televizní vysílání, počítačové a mobilní sítě

multimédia překódovat na kvalitu – datový tok – kvalitu, dle rychlosti dostupného připojení. Je možné data uložit vícekrát v různých formátech s různou kvalitou. Rozumnější někdy může být užití jediného formátu pro uskladnění všech typů médií, ze kterého se překóduje požadovaný výstup dle požadavku přehrávacích zařízení (3GP od 176 x 144 px, oproti HDTV až 1920 x 1080 px), odpovídající přenosové rychlosti a požadované kvality služeb QoS [Kos03].

Prvním (1996) a v současné době ve verzi 5 (2003) nejpoužívanějším real-time protokolem pro Voice over IP (VoIP), video a dokumenty je International Telecommunication Union (ITU) H.323. Pokrývá jak paketovou komunikaci (IP), tak přepínanou (ATM, telefonní síť). Je navržen pro univerzální video konference (point to point i multipoint) [ITU06].

Video kodekem je standard příbuzný MPEG-2, H.261 (1990), s rozlišením CIF (352 x 288 px) nebo QCIF (176 x 144 px), jeho datový tok se pohybuje v násobcích 64kbps. Novější verze je označena jako H.263 a H.264 (digitální vysílání HDTV), využívá MPEG-4. Audio kodeky jsou: G.711 s datovým tokem 64 kbps (PCM), G.723.1 s 5-6 kbps (ACELP) nebo GSM 06.10 s tokem 13 kbps (RPE-LTP). Kontrolní zprávy jsou zde zprostředkovány speciálním kanálem H.245.

Protokol H.323 je použit například sítěmi ISDN¹⁴, nebo mobilními sítěmi 3. generace (3GP, UMTS). Ilustrace je uvedena na následujícím obrázku Fig. 2.17.

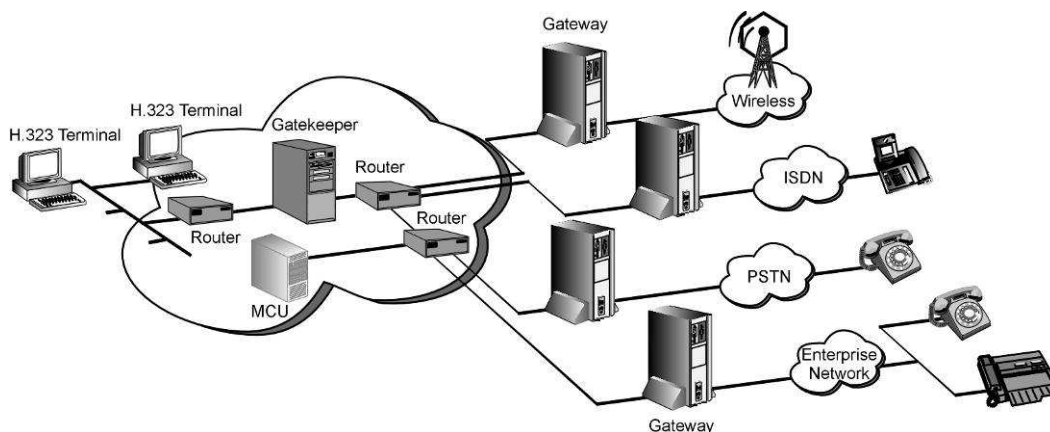


Fig. 2.17 Packet-based multimedia communications systems H.323 [RC06]

Jednodušší alternativou H.323 pro přenos hlasu (Voice over IP, VoIP) a videa je Session Initiation Protocol (SIP, RFC3261) [TIS02], skupiny Internet Engineering Task Force (IETF), která stojí za internetovými standardy IP, TCP, UDP, HTTP, SMTP a podobně. Firemní (Avaya) verze SIP je použitý například v telefonní síti VUT.

Existují i další IP protokoly pro real-time přenos MM dat, například Real-time Transport Protocol (RTP), který zajišťuje synchronizaci (ne QoS) pro streamování například MPEG-4. Používá se ve spojení s protokolem RTPS například pro streamování přednášek na FIT (rtps://video1.fit.vutbr.cz:5555/zaznamX).

Digitální televizní vysílání – DVB (Terestriální, Satelitní i Kabelové) je plně v režii MPEG-2. MPEG-1 a MPEG-2 interně obsahují obdobný předpis pro synchronizaci těchto (MPEG) dat. MPEG-4 využívá část 1. (systémy, ISO/IEC 14496-1), část 6. (MM distribuční integrační rámec) pro synchronizaci a část 8. pro přenos přes paketové síť. Pro HDTV digitální televizní vysílání se MPEG-4 používá vestavěně v H.264, protože se čeká na nový standard MPEG-21.

¹⁴ Integrated Services Digital Network, digitální přepínané telefonní služby

2.12 Prezentace

Prezentace je jednou z nejlépe vyřešených oblastí v MM, které technologicky předbíhá ostatní. Příkladem je schopnost libovolného kodeku MPEG-4 zobrazit 3D objekty, jejichž automatický popis v průběhu komprese zřejmě ještě neexistuje. Prezentace se obvykle rozděluje na audio a vizuální část, do které mohou spadat i metadata.

Pokud máme zajištěno médium, jsme schopni AV obsah s dostatečnou kvalitou doručit, navzájem synchronizovat, pak jeho přehrávání můžeme přenechat klientským zařízením (telefon, televize, přehrávače) nebo knihovnám a standardním aplikacím operačního systému nebo webového prohlížeče.

Jediným problémem může být, pokud daný systém nezná potřebný kodek nebo protokol. Problém řeší známé programy pro zobrazení a přehrávání, jako jsou <http://www.xnview.com/>, <http://www.winamp.com/>, <http://www.videolan.org/>, <http://www.real.com/>, <http://www.apple.com/quicktime/>, nebo jejich alternativy.

2.13 Bezpečnost a ochrana

Bezpečnost je nezbytným prvkem každého počítačového systému. Aplikace multimediálních systémů jsou citlivé například v oblastech zdravotnictví, průmyslu, vývoje a výzkumu nebo při tvorbě jiných autorských děl. Pro bezpečnostní MM systémy, například letiště, je tento bod stěžejním. Proto musí také MMDBMS zajistit ochranu dat způsobem, [Han00]:

- aby k nim měly přístup pouze oprávněné osoby (autentizace a autorizace),
- aby se zpracovávaly nefalšované informace (integrita a autenticita),
- aby se dalo zjistit, kdo je vytvořil, změnil nebo odstranil (nepopíratelnost),
- aby nebyly nekontrolovaným způsobem vyraženy (důvěrnost),
- aby byly dostupné vždy, když jsou potřeba (dostupnost).

Pod první bod spadá také ochrana intelektuální vlastnictví a nelegální kopírování obsahu. Pro některé osoby, společnosti a organizace jsou znalosti, patenty, nebo autorská práva jediným zdrojem příjmů.

Tyto vlastnosti MM databází zřejmě nelze garantovat obyčejným souborovým systémem, ale je nutné použít adekvátní MMDBMS.

Navíc je třeba dát pozor na sociální dopad výsledků z bezpečnostních AV dat (Velký bratr, 1984). To platí i pro návrháře, protože tato oblast je legislativně regulována – Zákon č. 101/2000 Sb., o ochraně osobních údajů, v aktuálním znění, Stanovisko č. 4/2004 ke zpracování osobních údajů prostředky kamerového sledování, Úmluva Rady Evropy č. 108/1981 o ochraně jednotlivců se zřetelem na automatizované zpracování osobních dat a směrnice 95/46/ES.

2.14 Architektury multimediálních systémů

Architektura libovolného systému je závislá především na konkrétní požadované aplikaci a je velmi podobná klasickým databázím, pokud jsou podobné požadavky. Ilustrace obecné architektury je uvedena na obrázku Fig. 4.1.

Může se jednat o obyčejný prohlížeč fotek na lokálním počítači a nepříliš vzdálených (optických) médiích. Obecně je ale možné říci, že se jedná o aplikaci typu klient-server. V nejjednodušším případě je klient tak tlustý, jak to jen jde – zprostředkovává všechny operace. Toto řešení ale není příliš univerzální a snažíme se přesunout co nejvíce operací na server obsluhující požadavky více klientů, lépe ale na systém řízení báze dat.

Proto je často médií a společnostmi poskytujícími MM obsah použita nejméně třívrstvá architektura systému – webový, aplikační nebo streaming server (front end) a MM DBMS – pro správu metadat a dat, které mohou být uloženy i na file serveru (FTP). Klienti této architektury jsou poté producenti i konzumenti multimedií. Příkladem je Oracle *interMedia* na obrázku Fig. 4.2.

V poslední době je nicméně začínají vytlačovat distribuované systémy, s mnoha zdroji a poskytovateli obsahu, založené například na MPEG-21, které umožňují přístup mnoha zařízením (televize, mobilní telefony apod. viz Fig. 2.17) k mnoha zdrojům včetně nějakých neexplicitních informací, získaných například z bezpečnostních kamer nebo medicínských snímků.

2.15 Aplikace získávání znalostí

Získávání znalostí¹⁵ je proces, jehož vstupem jsou běžná (AV) data a výstupem nové, zajímavé a potenciálně užitečné informace. Získávání znalostí je interaktivní, což znamená, že si uživatel vybírá požadované znalosti podmínky, při jejichž změně se proces opakuje. Je tedy iterativní. Proces se skládá z několika částí, jako čištění, integrace a transformaci dat, což zahrnuje extrakci rysů, ale především dolování dat a jejich následnou evaluaci a prezentaci.

Dolování dat¹⁶ je souhrnný název skupiny metod založených na umělé inteligenci, které data klasifikují, shlukují, objevují zajímavé vlastnosti, vztahy, změny a vybírají z nich jen to podstatné, co nás zajímá nebo překvapí.

Důležitou oblastí získávání znalostí je nalezení vztahu mezi syntaxí (černý flek) a sémantikou (pes). Z tohoto pohledu se dělí dle rovin aplikace:

- Obecné aplikace se snaží získávat znalosti ze vstupních dat, o kterých nemají žádnou apriorní znalost a nejsou schopny odvodit mnoho o sémantice.
- Speciální aplikace využívají účelových zdrojových dat, o kterých je známo kde byly získány a stanovit, jaké znalosti lze dolovat zvoleným přístupem. Je možné alespoň přibližně sestavit znalostní bázi, použitelnou algoritmy pro učení (jak vypadá pes).

2.15.1 Klasifikace

Klasifikace je proces, který hledá modely, popisující a odlišující známé třídy dat, tak aby bylo možno zařadit i neznámé objekty. Z tohoto důvodu není použitelná pro všechny roviny aplikace, ale pouze pro ty, kde jsme schopni určit znalostní bázi v podobě tříd.

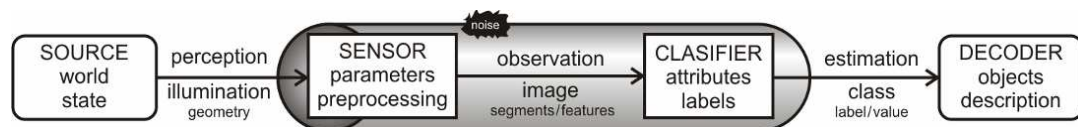


Fig. 2.18 Kanál klasifikace v počítačovém vidění [Chm006]

Pro ilustraci se nejlépe hodí naivní Bayesovská klasifikace, odvozená od počítačového vidění. Její proces je ilustrován na obrázku Fig. 2.19, formalizujícím Fig. 2.18. Zde y je skrytý stav, který si přejeme zjistit. Jediná informace, se kterou o daném objektu klasifikátor disponuje, jsou naměřená x , označovaná jako pozorování $\mu(w)$, které je ale zatíženo chybou. Obecně platí:

$$\text{Eq. 2.2} \quad \text{kompletní informace } (w) = \text{pozorovatelná } (\hat{w}) + \text{ztracená informace } (\text{šum } \tilde{w}).$$

¹⁵ Knowledge discovery in databases, KDD.

¹⁶ Data Mining, DM. Přesněji by se měl nazývat dobývání (informací) v datech.

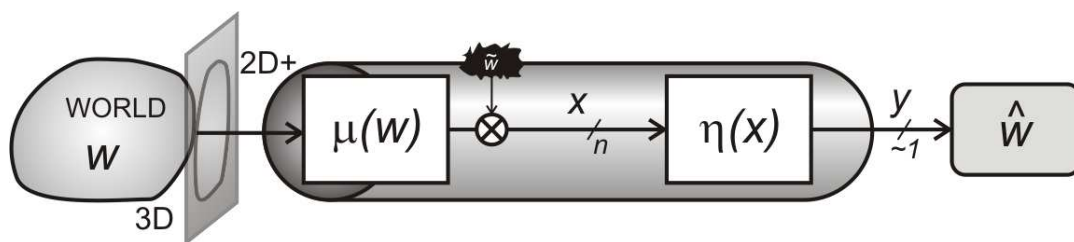


Fig. 2.19 Formální ilustrace klasifikace [Chm006]

Klasifikační funkce $\eta(x)$ obsahuje bázi dat, na základě které se rozhodne pro zařazení neznámého objektu, což je ilustrováno na následujícím příkladě.

Ex. 2.5 Příklad klasifikace neznámého hnědého, kulatého objektu

Tvar (x_1)	Barva (x_2)	Třída (y)	$P(y)$	$P(x_1/y)$	$P(x_2/y)$
<i>čtverec</i>	<i>hnědá</i>	<i>Čokoláda</i>	0,33	0,90	0,90
<i>kruh</i>	<i>červená</i>	<i>Jablko</i>	0,33	0,45	0,70
<i>kruh</i>	<i>oranžová</i>	<i>Pomeranč</i>	0,33	0,45	0,70
?	?	?	0,01	0,10	0,10

Před možností klasifikovat, neboli zařadit neznámý objekt do odpovídající třídy, je nutné znát charakteristické vlastnosti objektů této třídy. To je v případě *Čokolády* její *čtvercový* tvar a *hnědá* barva. V případě naivní Bayesovské klasifikace je proces učení jednoduchá statistika – *Čokoláda* se v trénovacích datech vyskytuje v 33% případů. To se označuje jako apriorní pravděpodobnosti výskytu objektu, $P(y)$. Při trénování se mohou vyskytovat další objekty, které nelze (nebo není radno) klasifikovat – (?).

Podmíněná pravděpodobnost $P(x_1/y)$ říká, že se v případě pozorování *čtverce* jedná o *Čokoládu* s 90% pravděpodobností. Tato pravděpodobnost se nazývá likelihood.

Naivní Bayesovská klasifikace, ve variantě maximalizace posteriorní pravděpodobnosti (Maximum A Posteriori, MAP), se snaží maximalizovat pravděpodobnost výskytu dané třídy pomocí její likelihood a pravděpodobnosti výskytu prostým násobením:

$$\text{Eq. 2.3} \quad \eta^*(x) = \arg \max_x P(x|y)P(y)$$

Tedy v našem příkladě se pokusíme určit třídu *hnědého kruhu* objektu pomocí MAP:

$$\begin{aligned} P(\text{kruh} | \text{Pomeranč}) \cdot P(\text{hnědá} | \text{Pomeranč}) \cdot P(\text{Pomeranč}) &= 0,45 \cdot 0,10 \cdot 0,33 = \\ &= 0,0149 \\ P(\text{kruh} | \text{Jablko}) \cdot P(\text{hnědá} | \text{Jablko}) \cdot P(\text{Jablko}) &= 0,45 \cdot 0,10 \cdot 0,33 = 0,0149 \\ P(\text{kruh} | \text{Čokoláda}) \cdot P(\text{hnědá} | \text{Čokoláda}) \cdot P(\text{Čokoláda}) &= 0,1 \cdot 0,9 \cdot 0,33 \\ &= \underline{0,0297} \end{aligned}$$

Maximalizací výsledku zjistíme, že se jedná o kulatou *Čokoládu*.

Bayesovská klasifikace se pro svoji jednoduchou interpretaci využívá také v teorii řízení. Její regresní (odhad spojité hodnoty) varianta je ale výpočetně složitější a teoreticky se neobejde bez složité mnohorozměrné integrace [Chm006]. Ta se v praxi nicméně obchází – vypočítá se střední hodnota a směrodatná odchylka diskretizované veličiny.

Pro klasifikaci a regresi existuje množství modifikací, které neobsahují naivní předpoklad, například Gaussovské, pracující s Eukleidovskou vzdáleností a

Gaussovským (rovnoměrným) rozložením pravděpodobnosti, původně využívané pro pozorování hvězd [Sor70].

SVM Obecným algoritmem pro tyto aplikace jsou kernelové metody, například Support Vector Machine. SVM se snaží generalizovat data a lineárně separovat jednotlivé třídy (bodů v mnohorozměrném prostoru – vektorů) od sebe separační hyper-rovinou Fig. 2.20, kterou je možné representovat skalárním součinem¹⁷ vektorů:

$$\text{Eq. 2.4} \quad \langle w, x \rangle + b = 0, \text{ pak klasifikace lze vyjádřit jako } f(x) = \sum \alpha_i y_i \langle x_i, x \rangle + b$$

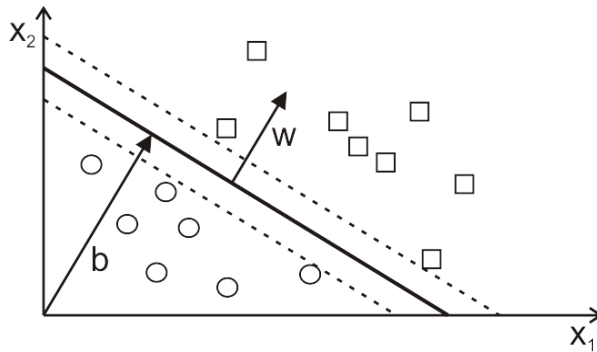


Fig. 2.20 Ilustrace lineární separace tříd pomocí SVM

V rovnici Eq. 2.4 jsou jednotlivé třídy y (čtverečky, kolečka) reprezentovány hodnotami $\{-1, 1\}$. Vektory nejbližší separační linii se nazývají support vectors, platí $\alpha = 1$.

Nalezení rovnice rozhodovací funkce, která maximalizuje vzdálenost mezi třídami, je závislá pouze na skalárním součinu vektorů, nikoli jejich poloze [COV95]. To je velice výhodná vlastnost, pokud není problém lineárně separovatelný.

V případě, že nelze třídy dat separovat, je nutné použít transformační funkci, zvanou kernel. Kernel K je skalární součin vektorů, které jsou transformovány funkcí $\phi(y)$ tak, aby byly lineárně separovatelné, jak je naznačeno v následující rovnici a obrázku.

$$\text{Eq. 2.5} \quad K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle, \text{ pak klasifikace lze vyjádřit jako } f(x) = \sum \alpha_i K(x_i, x)$$

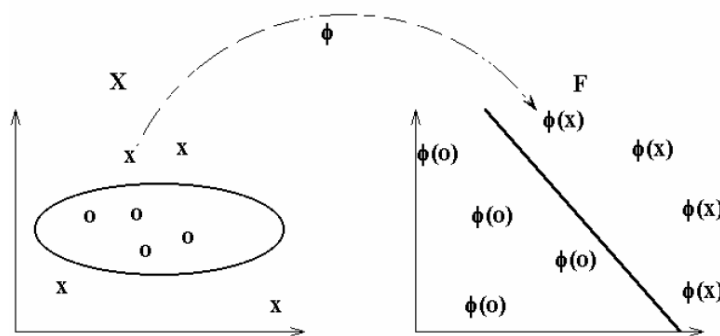


Fig. 2.21 Ilustrace transformace v kernelu pro SVM [predelat]

Příklady kernelů jsou například polynomiální funkce a RBF (Radial Basis Function):

$$\text{Eq. 2.6} \quad K(x, z) = \langle x, z \rangle^n, \quad K(x, z) = e^{-\|x-z\|^2/2\sigma}$$

¹⁷ anglicky inner product, vektorů a, b : $\langle a, b \rangle = a^T b = \sum a_i b_i$ (v ortogonální bázi)

Učení SVM a další více či méně známé algoritmy jako rozhodovací stromy, hrubé množiny, fuzzy logika, Markovovy skryté modely (Markov Hidden Models, HMM) nebo Kalmanův filtr, je možné najít v [Fis06].

Vhodné je také využití neuronových sítí [Zbo05].

2.15.2 Shlukování a analýza

Shlukování (segmentace) je technika, která umožňuje seskupit podobné objekty dle míry jejich podobnosti, kterou je možné měřit za stejných podmínek jako v kapitole 2.8.1. Shlukování se snaží data zařadit do skupin, které předem nezná, respektive je vytvořit.

To, co brání shlukování v její naprosté autonomii, jsou dvě věci. První z nich je počet tříd shluků, který je možné stanovit i z maximální možné vzdálenosti dvou shluků, ale ta už je třeba nastavit manuálně, dle požadované aplikace. Zásadním problémem je dále inicializace těchto shluků. Metoda K-Means [Zbo04] určuje jejich prvotní rozmístění náhodně, což není ideální řešení, ale někdy se může podařit. Pro vizuální data bych raději volil například velikost gradientu (detekce hran) – místa s velkým gradientem mohou oddělovat oblasti. Obdobně i místa s nízkou hustotou prvků pro jiný typ dat.

Asociační analýza se snaží nalézt vztahy mezi analyzovanými daty. Ty se vyjadřují asociačními pravidly tvaru implikace $A \Rightarrow B$. Společně s metrikami podpora $S(A \Rightarrow B) = P(A \cup B)$ a spolehlivost $C(A \Rightarrow B) = P(B / A)$ určují, které hodnoty atributů se vyskytují v dané množině dat s jistou pravděpodobností společně. Více v [Han01].

Dolování asociačních pravidel z multimedií lze chápat jako automatizovanou klasifikaci se stanovenou mírou zajímavosti. Každý objekt lze považovat za transakci pro nalezení frekventovaně se nacházejících vzorů dat, nahrazujících položky. Zaměřím se na vizuální data:

- Vztahy mezi obsahem a popisem multimédia jsou pravidla typu: „Pokud je značná část horní poloviny modrá, zřejmě se jedná o oblohu.“ Tato kategorie je sblížením obsahově a popisově zaměřeného získávání informací, protože z obsahu obrázku nebo scény usuzuje na klíčové slovo a tím je obloha. Nebo naopak při klíčovém slově „ne“ lze z videa vypořadovat mírný pohyb hlavou.
- Asociace obsahu dat bez určení jejich umístění – do této kategorie náleží například výroky jako: „Pokud obrázek obsahuje dva modré čtverce, pak je zde obvykle i žlutý kruh.“ Protože se jedná o popis obsahu obrázku, ale žádný objekt zde není lokalizován.
- Asociace obsahu dat se specifikací jejich lokalizace: „Pokud je červený trojúhelník mezi dvěma modrými čtverci, pak jsou všechny umístěny ve žlutém kruhu,“ náleží k této kategorii, objekty jsou asociovány pomocí jejich prostorových vztahů.

Evoluční analýza popisuje „zvyky“ objektů, jejich změnu v průběhu času a snaží se odhalit jejich trendy. Okrajová analýza se naopak snaží hledat nezvyklé vzory v chování objektů, které jiné metody označí jako šum. Používají se metriky pro měření odchylky od tříd, například pro odhalení nekalého záměr nebo činu.

Získávání znalostí demonstrují prostřednictvím následujícího příkladu.

Ex. 2.6 Příklad získávání znalostí aplikované na fotbalový zápas [Chm06]

Získávání znalostí z videozáznamu fotbalového utkání je poměrně speciální záležitost, do jisté míry je možné popsat fotbalový zápas jednoduchými pravidly, kterým je schopen porozumět jak fanoušek Baníku, tak i počítač. Kamery zabírají reálnou zelenou plochu, která je rozdělena a ohraničena bílými liniemi. Mimo ně nás záznam obvykle nezajímá. Na hřišti je několik typů objektů – hráči (2x10 odlišné barvy), statické branky

s brankáři libovolné (šedé) barvy a míč se specifickými vlastnostmi. Rozhodčí, případně ostatní postavy klasifikovat nebudeme. Ústřední událostí je míč v brance. Předmětem získávání znalostí z multimediálních a obecně časoprostorových databází je například to, kdo se snaží dostat míč kam, jakým způsobem se mu to daří, kdo přihrává komu, jak dlouho drží míč, kde a vyvodit z toho jisté závěry.

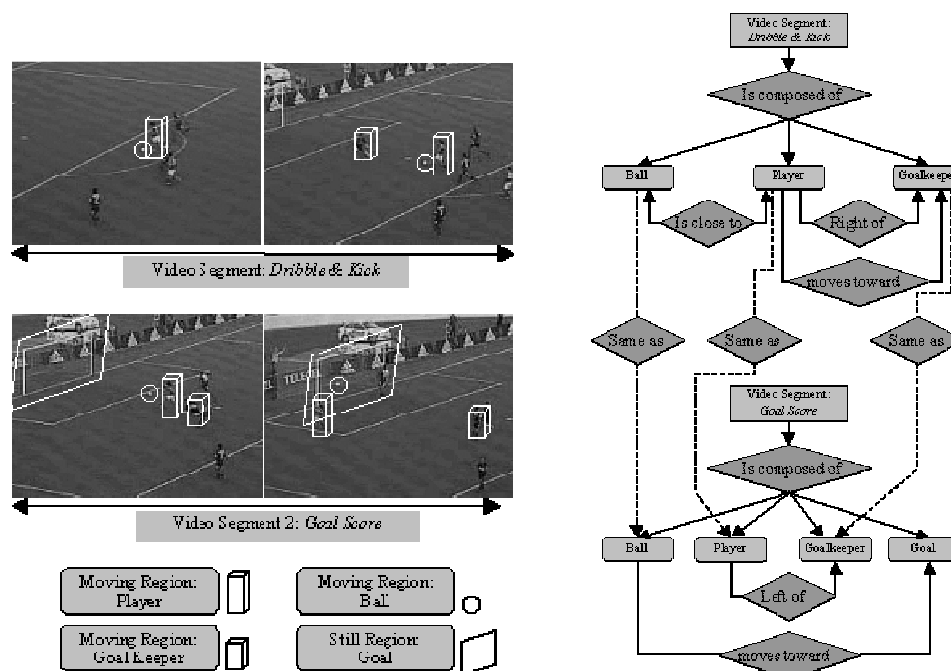


Fig. 2.22 Příklad rozpoznání objektů videa a jeho reprezentace z [M704]

„Jak lze dosáhnout popisu jako na obrázku Fig. 2.22?“ Extrakci těchto rysů lze rozložit do několika částí:

- Definice zdrojů a cílů, je nutné vytvořit znalostní bázi obsahující syntaxi (jaké objekty se v obraze mohou vyskytovat, jak vypadají, vytvořit jeho modely) a navázat na ně sémantiku (co daný objekt reprezentuje) informací, které se v obraze mohou vyskytovat. Příkladem je popis o odstavce výše.
- Extrakce objektů – každý pohyblivý region může být označen za objekt a/nebo rozdělen do podmnožin, které budou analyzovány samostatně, ať už se týká vektoru pohybu (prolínání hráčů, ruce, nohy) nebo jejich popisu jako statických obrázků. MPEG-7 poté nabízí rámec pro uložení těchto časoprostorových dat a jejich vzájemné závislosti.
- Transformace do souřadnic reálného světa – definice a vytvoření snímaného prostoru, specifikace polohy a pozice senzorů (kamer) a transformace detekovaných objektů v obraze do reálných souřadnic za pomoci stacionárních bodů na hřišti (lajny, branky).
- Vytvoření databáze pohybujících se objektů na různých úrovních – dle MPEG-7, ale je možné vytvořit i časoprostorovou databázi a do ní ukládat pohyby jednotlivých objektů tak, aby z nich bylo možné snadno vyvodit další informace.

Z této databáze lze vyvozovat znalosti následujícími přístupy:

- Klasifikace a predikce. Může sloužit na jedné úrovni pro estimaci pohybu a jeho vyhodnocení například v okamžiku, kdy kamera nezabírá dané objekty, nebo na globální úrovni může odhadnout výsledek zápasu Most – Liberec na 0 : 2.

- Shlukování má obecně široké uplatnění, může například rozřazovat hráče podle jejich příslušnosti do týmů nebo skupin (útok, obrana). Na nižší úrovni slouží pro extrakci pohyblivých regionů.
- Asociační analýza dokáže například odhadnout, že pokud při zápasu Sparta – Arsenal nahraje míč Pirés (7) Henrymu (14), pak padne do soupeřovy branky s pravděpodobností 66%.
- Evoluční analýza popisuje obvyklé chování objektů a mohla by například být schopná odhadnout, komu nahraje daný hráč a s jakou pravděpodobností nebo odhadnout výsledek utkání podle první poloviny.
- Okrajová analýza by naopak měla být schopná rozpoznat různé nestandardní situace, jako je zranění hráčů nebo změna taktiky.

Zamyšlení ke 2. kapitole

Jaká data se vyskytují v databázích?

Pro jaká data není vhodné použít relační databázi?

Jak je to s jejich strukturou? Představte si data například pro medicínu.

Znáte operace, které se provádějí na těchto datech?

Pro jaké účely jsou vhodné (multimediální) databáze?

Popište vyhledávání. Dle čeho je možné vyhledávat?

Co je nutné udělat před vyhledáváním?

Znáte popis multimediálních dat? Jaký existuje a jak lze získat?

Stručně shrňte požadavky na systém řízení báze multimediálních dat.

Jaká znáte multimediální data?

Jaké formáty se používají pro obraz, zvuk, video. Jaké pro dokumenty?

Co je to mediální kontejner?

Co jsou to metadata? Jaké typy a metadat znáte?

K čemu slouží MPEG-7? Co definuje?

Jaká média popisuje MPEG-7? Jak mohou být uložena?

K čemu slouží MPEG-21?

Co jsou to rysy, jak je lze získat a k čemu slouží?

Jaké rysy je možné extrahovat z audia, obrázků, videa nebo textu? Pro jaké aplikace?

Vysvětlete rozdíl ve vyhledávání dle obsahu a popisu.

Co je to podobnostní vyhledávání, vysvětlete jeho principy.

Co je to index, k čemu slouží a jaké techniky se využívají.

Pokuste se popsat možnosti ukládání a jejich výhody.

Jak je možné přenášet média? Jakým prostředím a protokolem?

Popište bezpečnostní opatření potřebná pro kritické aplikace a jiné zvláštnosti.

Zkuste načrtnout architektury multimediálních systémů.

Naznačte typy znalostí extrahovatelné z multimedií.

Jaký je rozdíl mezi klasifikací a segmentací?

Jaký je rozdíl mezi Bayesovskou klasifikací a SVM?

3 Databázové technologie a standardy

Následující kapitola se zabývá databázemi obsahujícími komplexní typy dat, jako jsou složité strukturované objekty, prostorová, temporální, hypertextová a především multimediální data a také technikami pro přístup k těmto databázím.

Relační model je založen na teorii množin. Entita reálného světa je reprezentována relací, podmnožinou kartézského součinu domén, množin atomických hodnot. Data jsou uložena jako n-tice hodnot těchto domén. Více v [Zen05].

Stávající rozšíření relačních systémů řízení báze dat (DBMS) dokazuje, že jsou pro většinu aplikací, především obchodních, stále plně dostačující. Nicméně pro složitější aplikace je tato „plochá“ organizace jednoduchých dat velmi omezující:

- Omezení první normální formou (atomičnost) nedovoluje odvozené, komplexní a vnořené (nested) struktury jako pole nebo množiny.
- Oddělená data a operace pro jejich manipulaci.
- Omezený počet vestavěných datových typů a složité dotazy SQL.

Z těchto důvodů je nutné adoptovat koncepty objektově orientovaného (OO) paradigmatu, které umožňuje lepší korespondenci mezi složité strukturovanými reálnými daty. To je vyžadováno v mnoha oblastech, jako je grafika, kartografie, medicína, fyzika nebo chemie [Cha01]. Hlavní předností objektových a objektově-relačních databází pro multimediální databázové systémy je možnost specifikovat komplexní datové typy a operace pro jejich manipulaci nezávislé na dotazovacím jazyce.

Objektový model dat v databázových systémech vychází ze známých principů objektově orientovaného modelování a programování. Objekty jsou použity pro modelování entit reálného světa. Mají svoji unikátní identitu a tvoří kolekce. Třídy jsou využívány pro definice množin objektů. Objekty téže třídy mají stejnou strukturu i operace (procedury a funkce se nazývají metody). Třídy jsou organizovány ve zájemně provázané hierarchii a nahrazují databázové schéma.

Objektově orientovaný přístup je však dále obohacen o techniky perzistence, reprezentace vztahů, dotazování, transakčního přístupu a tak podobně.

Persistentní znamená (objekt) existující bez časového omezení, do té doby než je úmyslně zničen. Úmyslné zničení je referenční nebo kaskádová akce (mazání, destrukce), nikoli ukončení transakce, sezení nebo aplikace [SQL03].

3.1 Objektově orientovaný koncept

Požadavek na perzistentní existenci komplexních datových typů a vztahů mezi nimi, jako rozšíření objektově orientovaného programování, byl zprvu řešen velkým množstvím vzájemně nekompatibilních řešení. To výrazně ovlivnilo jejich nasazení v praxi. Proto byla malými výrobci objektových databází založena skupina Object Database Management Group (ODMG), jejímž cílem bylo vyvinout nezávislý standard pro manipulaci a uchovávání objektů v databázích. Tato snaha vedla k výsledku v podobě knihy Object data standard: ODMG 3.0 [Cat00], v současné podobě třetí verze tohoto standardu. Skupina ODMG se v roce 2001 přidružila ke skupině Object Management Group (OMG), tvůrci architektury CORBA a správci UML.

Implementace je definována pro objektové programovací jazyky C++ a Smalltalk. Standard ODMG pro jazyk Java – Java Data Objects (JDO) je spravován společností Sun Microsystems. JDO je navrženo tak, aby bylo možné se věnovat objektovému modelování a programování a starosti s persistencí přenechat implementaci JDO.

Object Data Standard sestává ze tří částí – referenční objektový model (ODMG Object Model), definiční jazyk Object Definition Language (ODL) a dotazovací jazyk Object Query Language (OQL).

3.1.1 Objektový model ODMG

Specifikuje standardní model dat pro objektově orientované databáze. Je definován jako nadmnožina objektového modelu OMG (Object Management Group). V následujícím přehledu jde hlavně o popis perzistentního rozšíření pro objektové databáze.

Objekt je jednoznačně identifikovatelná entita reálného světa, má stav a chování. Stav sestává z hodnot vlastností – atributů a vazeb mezi objekty, chování je specifikováno pomocí operací (metod). Ne všechno v objektové databázi je objekt – literály (např. čísla) nemají, na rozdíl od objektů, jednoznačný Object Identifier (OID), jsou reprezentovány vlastní hodnotou [Cha01]. Množina objektů stejného typu – kolekce, také není objekt [Hru05]. Objektové databáze, na rozdíl od relačních, umožňují odkaz na základě identity OID. Z toho plyne také dvojitý pojem rovnosti – na základě identity a stavu – rekurzivní hodnoty vnořených atributů.

Data a operace nad nimi jsou zapouzdřeny do jediné struktury: objektu. Zapouzdření poskytuje logickou nezávislost dat a je důležité pro abstraktní třídy. Z tohoto pohledu se objekt skládá z (několika) rozhraní a (jedné) implementace. Ne všechny vlastnosti jsou viditelné i z vnějšku objektu, rozhraní zpřístupňují pouze ty veřejné (public). Chování objektu zajišťují metody, které sestávají z hlavičky, ta specifikuje jejich parametry a typ případného výsledku, včetně jejich implementace v libovolném výpočetně úplném OO programovacím jazyce. V objektových databázích tedy není dotazovací jazyk (OQL) nezbytný [Cha01].

Třída je šablona pro tvorbu objektů, specifikuje strukturu atributů a množinu metod. Typ objektu je obecnější pojem, reprezentovaný obvykle jménem třídy. Každý objekt je nějakého typu a má s ostatními objekty stejného typu stejnou množinu vlastností a operací. Objektové databáze poskytují množství předdefinovaných typů (např. pro geometrické obrazce), které je možné libovolně rozšířit. Organizace (vztahů) tříd v agregační hierarchii představuje schéma tříd, například v UML diagramu databáze.

Třídy mohou mít vlastní proměnné a operace, například počet instancí nebo operátor new(). Kolekce všech instancí generovaných jednou třídou se nazývá extant. Důležitou vlastností je druh jejich perzistence:

- Implicitní vlastnost všech instancí třídy. Objekt je automaticky uložen do databáze při jeho vytvoření. Tyto systémy mají tedy dva typy konstruktorů – pro vytváření perzistentních a temporálních instancí tříd.
- Vytvoření objektu nemá vliv na jeho perzistenci. Ta je zajištěna například vložením do perzistentní kolekce.

Obdobně je to s mazáním objektů a integritou:

- Persistentní objekt je smazán až po smazání všech referencí na něj. Například pomocí garbage collection.
- Objekt může být smazán explicitně a porušit tak referenční integritu. Ta musí být zajištěna na aplikační úrovni.

Persistentní objekty mají svoji jednoznačnou identitu OID po celou dobu existence, nemění se tedy ani v případě změny příslušnosti ke třídě. Evoluce objektu se nazývá migrace.

Dědičnost umožňuje odvodit definici stavu a chování (pod)typu z definice jiného (nad)typu. Jednoduchá dědičnost mezi třídami je používána jako rozšíření (extends).

Pokud je nadtypů více, mluví se o vícenásobné dědičnosti. Systémy, které ji podporují, musí umět řešit případné konflikty. ODMG umožňuje vícenásobnou dědičnost chování (ISA) pouze v případě, že se nedědí dvě stejně nazvané operace náležící různým typům.

Dědičnost znamená, že je možné použít (nahradit) instanci podtypu tam, kde je požadován nadtyp, čímž vytváří konzistentní hierarchii. Pro implementaci znamená výhodu sdílení a znovupoužitelnost kódu mezi třídami. Operace mají stejné jméno, ale pro každou třídu je možné vytvořit specifickou implementaci (overriding). Výběr polymorfní nebo typově přetížené (overloading) operace se provede za běhu programu (late binding). Dědičnost má dále za následek klasifikační hierarchii mezi extenty jednotlivých tříd a využívá se pro modelování jako specializace a generalizace.

3.1.2 Definiční jazyk ODL

ODL definuje objektové databázové schéma odpovídající objektovému modelu. Součástí této části je i jazyk pro výměnu objektů a jejich zálohování na disk Object Interchange Format (OIF), ale na jeho místě se ujal univerzální jazyk XML.

Namísto referenčního přehledu ODL [Cat00] uvedu příklad multimediální objektové databáze pro řízení projektů. Její schéma tříd je uvedeno na následujícím obrázku.

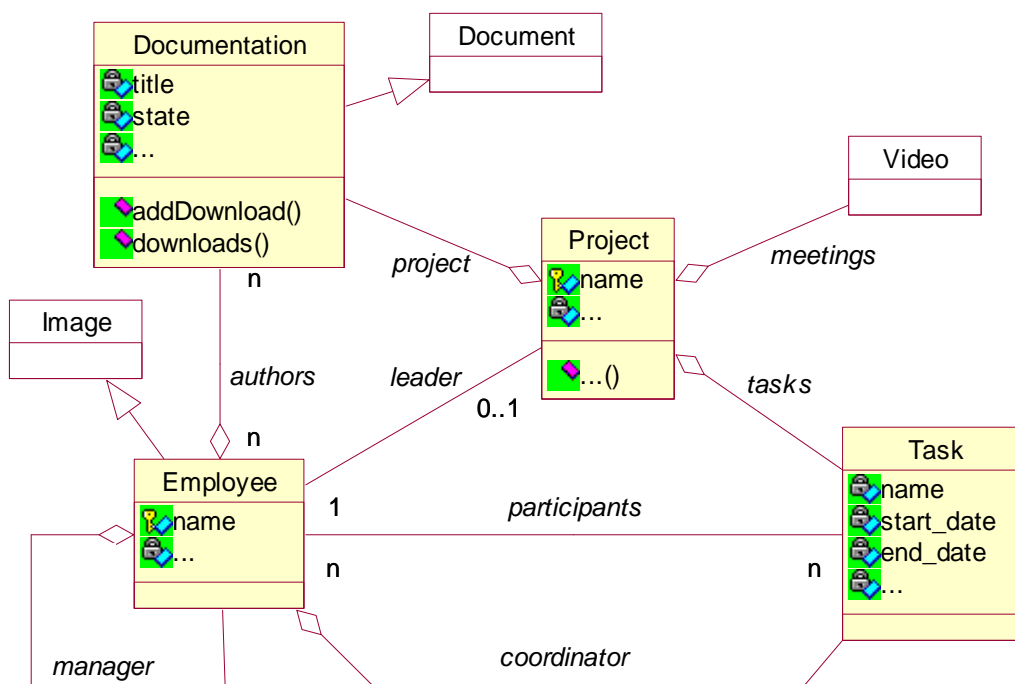


Fig. 3.1 Diagram tříd multimediální aplikace pro řízení projektů (pouze ilustraci)

Schéma databáze na obrázku Fig. 3.1 (nenapodobovat!) obsahuje 4 zjednodušené uživatelské třídy a 3 předdefinované systémem. Jsou to třídy pro uchování obrázků, videa a dokumentů. Třída zaměstnanců je odvozena od třídy obrázků pro uchování jejich fotografií, nabízí metody pro zobrazení, načtení ze souboru libovolného typu, nebo dotazování dle obsahu obrázku. To se nemusí v tomto kontextu zdát důležité, ale technika pro rozpoznání obličeje může sloužit pro kontrolu přístupu, prezence zaměstnanců na poradách, případně zabránit tomu, aby vedoucí oddělení zaměstnával pouze své příbuzné (ale už ne milenky).

Obdobně je to s třídou dokumentací, která rozšiřuje standardní třídu Document o některé atributy, například autorství a funkci která zvyšuje počítadlo při jeho stažení. Naopak

video záznam porad není nutné rozšiřovat, protože metadata videa spolu s přepisem meetingu pomocí nějakého systému pro rozpoznání řeči je dostatečné pro jeho identifikaci.

V diagramu se náhodně vyskytují dva typy vazeb. První z nich je vztah asociace (relationship), který je obecně neorientovaný a zajišťuje referenční integritu. Pokud chceme explicitně vyjádřit orientaci vztahu, je možné použít orientovanou asociaci. V našem příkladě používám raději agregaci, která obecně neimplikuje vlastnictví a nezajišťuje integritu. To odpovídá vazbě použitím attribute v následujícím kódu.

Ex. 3.1 DDL multimediální aplikace pro řízení projektů

```

class Employee EXTENDS Image
( extent Employees
  key name )
{
  attribute string name;
  attribute Employee manager;

  relationship Project leads
    inverse Project::leader;
  relationship Set<Task> participate
    inverse Task::participants;
}

class Project
( extent Projects
  key name )
{
  attribute string name;
  attribute List<Document> documents;
  attribute Set<Task> tasks;
  attribute Set<Video> meetings;

  relationship Employee leader
    inverse Employee::leads;
}

class Task
( extent Tasks )
{
  attribute string name;
  attribute date start_date;
  attribute date end_date;
  attribute Employee coordinator;

  relationship Set<Employee> participants
    inverse Task::participates;
}

class Documentation EXTENDS Document
( extent Documentations
  key title )
{
  attribute string title;
  attribute string state;
  attribute int downloads;
  attribute Project project;
  attribute List<Employee> authors;

  void addDownload();
}

```

```
int downloads();
}
```

Definiční jazyk může specifikovat mnoho dalších konceptů [Cat00]. Například rozhraní, v tomto příkladě zpřístupňující třídu zaměstnance bez jeho fotografie.

3.1.3 Dotazovací jazyk OQL

OQL je deklarativní (neprocedurální) dotazovací jazyk silně inspirovaný SQL. V relačních databázích je dotazovací jazyk jedinou možností pro přístup k datům. Objektově orientované aplikace používají spíše navigační způsob – navigace po identifikátorech objektů v agregační hierarchii tříd. Oba způsoby se ale doplňují. Dotazem je možné získat například množinu dat, se kterými dále manipulujeme navigačním způsobem. To je výhodné například pro algoritmy využívající procházení grafem (backtracking), protože ušetří velké množství náročných operací spojení (join).

Objektový dotazovací jazyk doplňuje relační o možnost spojení implicitně, pomocí hierarchické struktury objektů. Pokud vezmeme v úvahu schéma databáze z předchozí kapitoly, můžeme položit dotaz s explicitním spojením:

Ex. 3.2 Dotaz OQL s explicitním spojením

```
select d
from Documentations d, Projects p
where d.project = p.oid and
      p.name = 'Nošení dříví do lesa';
```

Nehledě na fakt, že systém nemusí poskytovat přístup k vestavěnému identifikátoru objektu, lze tento dotaz napsat mnohem úsporněji, pomocí implicitního spojení:

Ex. 3.3 Dotaz OQL s implicitním spojením

```
select d
from d in Documentations
where d.project.name = 'Vynález kola';
```

Tak, jak SQL umožňuje vytvářet libovolné nové relace (pohledy) výsledku dotazu, umí OQL vytvářet nové struktury. Samozřejmostí je také možnost volání funkcí v dotazech:

Ex. 3.4 Dotaz OQL s implicitním spojením

```
select distinct struct(nazev: d.name, stahnuto: d.downloads)
from Documentations d
where d.downloads > 0;
```

Výsledkem tohoto dotazu je množina n-tic **Set<struct(nazev: d.name, stahnuto: d.downloads)>**. Nutno poznamenat, že takto vytvořené struktury mají novou identitu a v tomto případě také ztratí informace o agregační hierarchii v databázi. Proto je využití struktur definovaných až při dotazu pro další dotazy značně omezené. OQL samozřejmě vyhodnocuje Booleovské predikáty pro objekty vyhovující zadanému dotazu, agregační funkce a podobně.

Jazyk OQL má oproti SQL několik dalších vylepšení, například také možnost dotazu na třídu, nebo i všechny její podtřídy. Ale jeho reálné implementace nebývají dokonalé a plná podpora vlastností specifikovaných ODMG často chybí nebo se liší [Cai03].

3.2 Objektově relační koncept

Začátkem 90. let si i velcí výrobci databázových systémů uvědomili nezbytnost objektového přístupu. Nicméně zjistili, že z obou přístupů je možné vybrat ty lepší vlastnosti. Z OO konceptu zapouzdření, dědičnost a polymorfismus. Relační koncept přináší řešení persistence, integrity, řízení (souběžného) přístupu a zotavení po chybě.

Objektově relační (OR) model představuje alternativu čistě objektovému. Je, na rozdíl od OO modelu, spíše evolucí než revolucí DBMS. Integruje objektové přístupy do relačního modelu, umožňuje tak koexistenci obou typů dat a je zpětně kompatibilní. To mu zaručilo přízeň konzervativních zákazníků – přechod z relačního DBMS na objektově relační systém řízení báze dat (ORDBMS) nevyžaduje velké změny ani investice. Naopak použití OR přístupu může usnadnit vývoj aplikací i v těch oblastech, kde nejsou persistentní objekty nezbytně vyžadovány, protože reference mezi objekty výrazně usnadňují dotazování. V neposlední řadě je mapování objektových modelovacích technik (UML) výrazně přirozenější, než metody převodu diagramu tříd na relační model.

Koncem 90. let OO a OR přístup v databázových systémech do značné míry konvergoval. Například jazyk OQL je značně podobný SQL:1999. Podstatný rozdíl je jen v přístupu k objektovým datům:

- OODBMS lze chápat jako rozšíření programovacího jazyka – perzistentní objekty jsou vytvořeny v prostředí objektového jazyka (Java, C++, Smalltalk) pomocí tříd definovaných v těchto jazycích a manipulovány navigačním způsobem. Jsou tedy silně provázané (pouze objektové paradigma) a flexibilní pro aplikace na straně klienta – tyto aplikace dočasně uchovávají všechna potřebná data.
- ORDBMS oproti tomu poskytují pouze aplikační rozhraní, založené na SQL, pro manipulaci s daty, jejichž definice typů (tříd) musí být DBMS známa předem. To poskytuje možnost umístění aplikace i na straně serveru se vším co s tím souvisí, jako je možnost velkého množství současně probíhajících nezávislých dotazů a jejich případná optimalizace.

Objektových rozšíření relačních databází vzniklo mnoho, proto se podíváme na standardy, které jejich vývoj usměrňují.

3.2.1 SQL:1999

Původním záměrem SQL:1999 bylo modifikovat SQL-92 tak, aby se z SQL stal jazyk objektově relačních databází [Eis99]. Sedmiletý vývoj ale celý dokument restrukturalizoval a přidal několik základních vylepšení. K těm patří nové datové typy:

- Large Object (LOB) v binární (BLOB) a znakové podobě (CLOB). Tyto typy jsou určeny pro zpracování velkého množství dat na straně klienta, proto má jejich využití v dotazech několik omezení oproti typům z SQL-92.
- Booleovský typ (BOOLEAN) s hodnotami true, false a unknown.
- Dva složené datové typy (ARRAY, ROW). Pole je indexovaná multimnožina prvků stejného typu (konečná kolekce). Složený atribut ROW lze chápat jako běžný řádek relační tabulky, proto neporušuje 1. normální formu. Pro jeho zpřístupnění se používá tečková notace, jako **Employees.Name.title = 'Ing.'** v následujícím příkladě.

Ex. 3.5 Alternativní definice tabulky užitím ROW dle SQL:1999

```
CREATE TABLE Employees (  
  empID INTEGER PRIMARY KEY,  
  Name ROW (
```

```

        title VARCHAR(5),
        first VARCHAR(30),
        mid VARCHAR(30),
        second VARCHAR(30),
        scitle VARCHAR(5),
    ),
    freeDays INTEGER ARRAY[7],
    inOffice BOOLEAN,
    photo BLOB
);

```

SQL:1999 přidává několik predikátů (např. **Employees.Name.title LIKE¹⁸ '%Prof%'**), zavádí rekurzivní dotazy, trigger, uživatelské role a vylepšuje transakční zpracování (ROLLBACK TO SAVEPOINT). Nejdůležitější je ale objektivě orientované rozšíření.

Informace o objektech nesou v ORDBMS strukturované datové typy (User-Defined Types, UDT). Ty mají následující vlastnosti:

- Skládají se z jednoho nebo více typů, jak vestavěných (INTEGER, ARRAY), tak dále strukturovaných. Ty se mohou libovolně zanořovat. Viditelnost atributů (private, public) není v SQL:1999 řešena.
- Jsou organizovány do hierarchií. Více specializované podtypy dědí z nadtypů všechny atributy i metody a mohou být doplněny o další. Jedná se o jednoduchou dědičnost. Je ale nutné explicitně zadat, že typ může mít podtyp (NOT FINAL). Také je možné vytvořit ekvivalent abstraktní třídy (NOT INSTANTIABLE).
- Atributy jsou zapouzdřeny pomocí systémem generovaných funkcí pro čtení a zápis (set, get). Tyto funkce nemohou být přetíženy (overloading), všechny ostatní ano.
- Jejich chování je zprostředkováno pomocí metod a funkcí. Metody jsou, na rozdíl od funkcí, vázány k danému typu. Funkce mohou být přetíženy, výběr se provádí v čase kompilace. Metody mohou být polymorfní, jsou vybírány dynamicky za běhu programu (late binding).
- Porovnání hodnot objektů se děje výhradně pomocí uživatelských funkcí.

Ex. 3.6 Definice typu zaměstnance odvozeného z typu obecné osoby a typované tabulky

```

CREATE TYPE tEmployee
  UNDER tPerson
  AS ( empID INTEGER,
       manager REF(tEmployee),
       salary REAL
       photo tImage )
  INSTANTIABLE
  NOT FINAL
  REF ( empID )
  INSTANCE METHOD
    moreMoney ( amount REAL )
    RETURNS REAL
);
-- typed table Employees
CREATE TABLE Employees OF tEmployee;

```

Přístup k atributům a funkcím do strukturovaných typů se provádí pomocí tečkové notace (**Employees.salary > 10000**). Přístup k atributům je možný i funkční notací.

¹⁸ Je možné použít také SIMILAR s regulárními výrazy

Reprezentace třídy (například v Javě) je možné vytvořit několika způsoby:

- Libovolný objekt se po standardní serializaci vlepí do BLOBu.
- Jednoduchý objekt je reprezentován řádem tabulky s vestavěnými typy.
- Objekt tvoří řádek s vestavěnými i strukturovanými typy, případně jediným UDT.
- Typovanou tabulkou (viz Ex. 3.6). Každému řádku je tedy přiřazen jeden objekt. Takovýto objekt je možné jednoznačně identifikovat elementem, který se chová jako OID. Jedná se o speciální typ reference (REF). Jeho použití je jednoduché: **Employees.manager->photo**.

3.2.2 SQL:2003

V současné době nejnovější verze standardu SQL [SQL03] přináší revizi všech částí standardu SQL:1999. V našem kontextu je nejdůležitější vylepšení objektově-relačního schématu, například o zavedení násobné dědičnosti. Dále klauzule WINDOW vhodné pro zpracování a dotazování souvislých proudů dat. Pokrokem jsou standardizované funkce pro podporu On-Line Analytical Processing (OLAP) [Hru05]. SQL:2003 dále standardizuje nativní podporu pro manipulaci s XML (SQL/XML), viz [Eis02].

SQL:2003 zavádí nové datové typy BIGINT, XML, ale především je to typ kolekce (MULTISET). Její použití je obdobné jako pole až na fakt, že tato multimnožina není uspořádaná ani teoreticky konečná. Deklaruje se třemi různými způsoby – **INTEGER MULTISET** může znamenat to stejné co **MULTISET[-1, -1, 2, 3, 4, 9]** nebo **MULTISET(SELECT i FROM t)**. Operace nad multimnožinou jsou podobné jako operace nad tabulkou včetně agregačních funkcí, ale existují mírné odchylky v názvech apod., viz [SQL03].

Nově byl přidán typ TABLE, který je interně definován jako multimnožina řádků (ROW). Tento typ je používán pro konverze výsledku funkcí na tabulku a opačně. Výhodou je, že tyto funkce mohou být definovány v jazycích jako C++ nebo Java. Je zde tedy možnost, jak definovat externí (EXTERN) tabulky. Jejich použití není složité, nicméně přesahuje, spolu s dalšími souvisejícími aspekty rozsah této práce [Eis04].

Užitečná možnost je deklarovat sloupec jako generovaný. To znamená automaticky při vložení nebo aktualizaci. Buďto bez možnosti přiřadit jinou hodnotu (GENERATED ALWAYS) nebo generovaný v případě, že nebyla hodnota zadána (GENERATED BY DEFAULT). Tato vlastnost lze použít například pro obecné funkce, OLAP, nebo pro definování identity řádku, jak je uvedeno v příkladech níže.

Ex. 3.7 Definice typu zaměstnance odvozeného z typu obecné osoby (jméno, ...)

```
CREATE TABLE Photos (
  photoID INTEGER GENERATED ALWAYS
  AS IDENTITY (
    START WITH 1
    INCREMENT BY 1
    MINVALUE 1
    NO MAXVALUE
    NO CYCLE ),
  name VARCHAR(100),
  desc CLOB,
  source tImage
)
```

V tomto příkladě je generovaný sloupec ilustrován jako sekvence (SEQUENCE). Ta představuje další novinku SQL:2003. Sekvenci je možné vytvořit samostatně a při vkládání hodnoty do tabulky ji zavolat pomocí NEXT VALUE FOR. Protože ale její

implementace, existovaly již dříve (AUTOINCREMENT v MySQL), existuje jednodušší možnost – použít sloupec s klíčovým slovem IDENTITY. Při vkládání pak není nutné hodnotu zadávat, je generována automaticky.

Další rozšíření umožňují jednodušší vytváření podobných tabulek v databázi (CREATE TABLE LIKE). Pokud požadujeme složitější operace jako změnu názvu a počtu sloupců, stejně jako pro kopii dat, je možné použít CREATE TABLE AS. Nicméně tyto funkce nevytváří logické vazby mezi daty. Pro jejich synchronizaci je možné použít 4. klauzuli pro modifikaci dat – sloučení (MERGE) .

SQL:2003 má oproti starším verzím mnoho dalších vylepšení [SQL03], jejich plná podpora ze strany tvůrců databází je nicméně, na rozdíl od ODMG, spíše otázkou času.

3.3 SQL / MM

SQL Multimedia (and Application Packages) měla původně (1999) přispět ve standardizaci abstraktních datových typů (ADT) a knihoven tříd pro vývoj a výzkum, textové databáze, zpracování dokumentů a správu multimediálních dat, jako jsou obrázky, zvuk, animace a video. Nicméně v průběhu standardizačního procesu (2001 – 2002) došlo k několika změnám a odklonu od obecných multimediálních dat, jak je vidět z následujícího přehledu částí SQL/MM:

Část 1 Framework: definuje společný koncept a předpisy.

Část 2 Full Text: zavádí typ FULLTEXT pro uchování dlouhých statí textu a některé další struktury, jako THESAURUS nebo SYNONYM. Definuje i vyhledávání odvozených slov ze stejného kmene pomocí STEMMED.

Část 3 Spatial: definuje prostorové typy a operace nad nimi.

Část 5 Still Image: zavádí dva základní typy. SI_StillImage pro uchování obrazových dat (Ex. 3.8) a SI_Feature pro uchování jejich rysů. Rysy mohou být SI_AverageColor (průměrná barva), SI_ColorHistogram (jstogram), SI_PositionalColor (rozložení barvy) a SI_Texture (texturní rysy). Všechny tyto vlastnosti obsahují metodu SI_Score pro výpočet podobnosti, která vrací hodnoty mezi 0 a 1 (Ex. 3.8). Jsou definovány i další operace nad obrázky, jako změna velikosti, ořezání nebo rotace.

Část 6 Data mining: definuje datové typy a operace pro klasifikaci, regresi, shlukování a vyhledávání (asociačních) pravidel.

Z multimediálních dat tedy zbyly pouze statické obrázky. Standard pro pohyblivé obrázky a zvuková data nebyl dokončen. Čtvrtá část měla definovat ADT pro obecné (matematické) účely, ale také nebyla dokončena.

Ex. 3.8 Datový typ SI_StillImage a podobnostní vyhledávání dle textury [Kuc04]

```
create type SI_StillImage as (  
  SI_content binary large object(SI_MaxContLength),  
  SI_contentLength integer,  
  SI_format character varying(8),  
  SI_height integer,  
  SI_width integer,  
  ...  
)  
  
SELECT *  
FROM Registrovana_Loga  
WHERE  
  SI_findTexture(Nase_logo).SI_Score(Logo) > 0.9;
```

Zamyšlení ke 3. kapitole

Po konzultaci

4 Komerční databázové systémy

První MMDBMS se začaly vyvíjet koncem 80. let. Patří mezi ně systémy ORION, ITASCA, ale první komerčně úspěšná, byla začátkem 90. let. především MediaDB společnosti MediaWay, která se v dnešní době jmenuje Chuckwalla. V současné době nabízí databázový systém pro vývojáře multimediálních aplikací na platformě Microsoft. Následující obrázek Fig. 4.1 zobrazuje její architekturu.

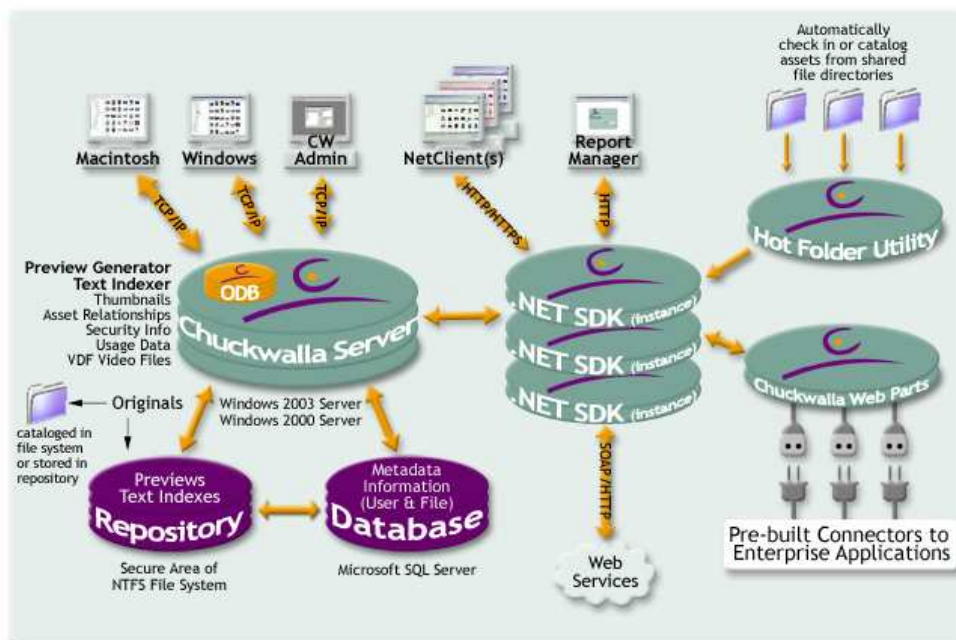


Fig. 4.1 Architektura multimediálního databázového systému (<http://www.chuckwallainc.com/>)

Dalším řešením disponuje například společnost IBM, která pro svoji databázi IBM DB2 Universal database vyvinula rozšíření (Extenders) pro audio, statické obrázky a video (<http://www-306.ibm.com/software/data/db2/extenders/aiv/>).

Nicméně zřejmě nejpropracovanějším řešením disponuje společnost Oracle.

4.1 Oracle interMedia

Oracle[®] *interMedia* je rozšíření databázového systému pro uchování, získávání, správu a dotazování obrázků, audia, videa, dokumentů a jim přidružených dat pro libovolné aplikace (www.oracle.com/pls/db102/portal.portal_db?selected=7).

Podporuje různé „nedatabázové“ operace (zpracování obrazu) a formáty pro uchování nestrukturovaných dat jako JPEG, MPEG, RA, QT, WMV, PDF, DOC nebo XLS. Pro ně nabízí jednoduché prostředky pro manipulaci s obsahem, například konverzi formátů, ořezání, změnu velikosti, nebo gama korekce. Pro textové dokumenty (výše uvedené včetně HTML a XML) je samozřejmostí efektivní indexace obsahu pro vyhledávání. Obdobně je tomu i pro obrázky. Možné je rozšíření o další speciální algoritmy pro zpracování multimedií, jako jsou další kodeky nebo subsystém pro rozpoznání řeči.

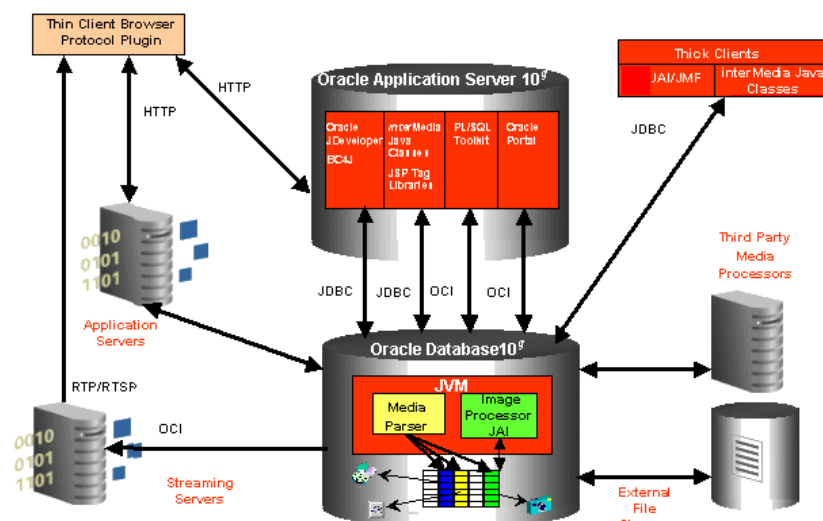


Fig. 4.2 Schéma Oracle interMedia (<http://www.oracle.com/technology/index.html>)

Oracle interMedia je založeno na objektově relačním DBMS. Obsahuje:

- Zdroj a jeho umístění (i mimo databázi) MM ve všech možných formátech.
- Formát MM a jiných, heterogenních, dat. Pro statické obrázky rozměry, rozlišení, počet barev, kompresní formát. U zvukových nahrávek použité kódování, počet kanálů, vzorkovací frekvence a doba trvání. Video má rámce popsané jako obrázek, jejich počet za sekundu, kompresní algoritmus, celkové trvání a datová propustnost, případně typ MIME.
- Metadata získaná z externích zdrojů, například autor, copyright, což jsou informace extrahované z jiných XML metadat a zpracovává zabudované informace, které jsou součástí MM jako je Exif, IPTC-IIM, XMP nebo medicínské DICOM.

Oracle ve verzi Enterprise obsahuje navíc nástroj pro popis statických obrázků dle jejich obsahu. Pro jejich indexaci a podobnostní vyhledávání. Vektor rysů (signature) obsahuje následující nízkourovňové elementy, které extrahuje při ukládání do DB:

- Barva, respektive její distribuce v obrázku – histogram.

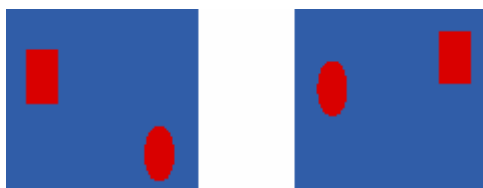


Fig. 4.3 Obrázky podobné dle barvy [Ora05]

- Textura charakterizuje zrnitost obrázku a hranovou frekvenci.

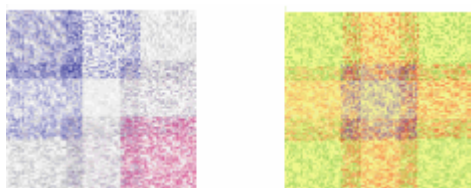


Fig. 4.4 Obrázky podobné dle textury [Ora05]

- Tvar objektů, získaných na základě segmentace obrazu, provedené podle barvy.



Fig. 4.5 Obrázky podobné dle tvaru [Ora05]

- Umístění barevných segmentů v obraze. Pomocí této vlastnosti se implicitně nevyhledává, ale slouží například k určení pozadí, jak ukazuje obrázek Fig. 4.7.



Fig. 4.6 Dva obrázky podobné dle barvy a umístění zároveň [Ora05]

Tento popis slouží pro podobnostní vyhledávání, realizované podle vážené vzdálenosti jednotlivých elementů v databázi, jejichž celková velikost je asi 3-4kB.

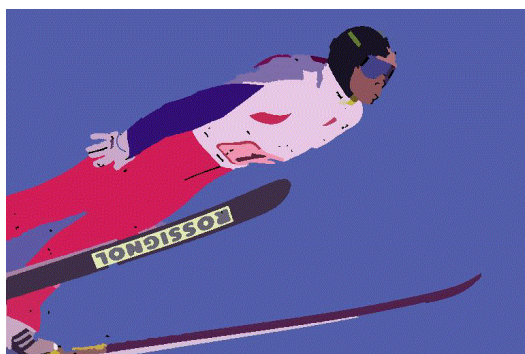


Fig. 4.7 Znázorňuje segmentaci do oblastí podle jejich barvy, z [Or05]

InterMedia aplikace je možné vyvíjet prostřednictvím jazyků Java, PL/SQL a některých dalších, příklady jejich použití jsou uvedeny v [Or05] a příloze 2 tohoto dokumentu.

Zamyšlení ke 4. kapitole

Jak by mohla vypadat nejjednodušší multimediální databáze? (Ve smyslu DVB přijímač s diskem a ethernet rozhraním nebo Galerie umění). Co taková databáze dokáže?

Jaké multimediální systémy řízení báze dat znáte?

Popište operace a data, nad kterými jsou prováděny.

Zkuste si implementaci MM databáze. Návod je uvedený v příloze 2.

5 Shrnutí

Tato práce představuje multimediální databáze jako celek. Je kompilací dostupné literatury, kterou mírně doplňuje (v oblastech, které se mi zdají zajímavé).

Namísto (nepříliš vydařené) úvodní kapitoly, je možné za úvod do problematiky multimediálních databází považovat kapitolu 2.2. Práce se zabývá jednotlivými požadavky multimediálních databází a jejich odchylkami od databází klasických, založených na relačním modelu a dalších systémů.

Práce se zaměřuje na popis obsahu založeném na primitivních vlastnostech multimédií, například barva, textura, tvar, pohyb u vizuálních, nebo frekvenční charakteristika u audio dat. Tyto vlastnosti jsou použitelné pro vyhledávání maximálně střední úrovně.

V kapitole o získávání znalostí (2.15) jsou naznačeny metody, které dokáží propojit nízkouúrovňové rysy s vysoce žádaným sémantickým popisem. Problémem je především obrovská diverzita obecných multimediálních dat, proto jsou aplikace pro nalezení informací obsažených v multimediálních datech úzce omezeny. V tomto vidím největší možnost rozvoje multimediálních databází.

Literatura

- [Bak87] BAKEN, R. J. *Clinical Measurement of Speech and Voice*. Taylor and Francis Ltd, 1987. 518 s. ISBN 0316078336.
- [Cai03] CAIXUE, Lin. *Object-Oriented Database Systems: A Survey* [online]. UCSC, 2003 [cit. 2006-05-25]. Dostupný z: <http://www.soe.ucsc.edu/~lcx/courses/cmcs277/cmcs277-project.pdf>.
- [Cat00] CATTELL, R. G. G. et al. *Object data standard: ODMG 3.0*. 2000...
- [Čer05] ČERNOCKÝ, Jan. *Stránka pro studenty kursu "Signály a systémy - ISS"* [online]. 2005 [cit. 2006-05-16]. Dostupný z: <http://www.fit.vutbr.cz/~cernocky/sig/>.
- [Cha01] CHAUDHRI, B. A. – ZICARI, R. *Succeeding with object databases*. J. Wiley, 2001. 442p. ISBN 0-471-38384-8.
- [Chl05] CHLUM, O – MATAS, J. Matching with PROSAC - progressive sample consensus. *Proc. of Conference on Computer Vision and Pattern Recognition* [online]. 2005. [cit. 2006-06-19]. Dostupný z: <http://cmp.felk.cvut.cz/~matas/>.
- [Chm006] CHMELAŘ, P. Bayesian Concepts for Human Tracking and Behavior Discovery, *Student EEICT 2006* [online]. VUT Brno, 2006. Dostupný z: http://www.feec.vutbr.cz/EEICT/2006/sbornik/03-Doktorske_projekty/07- Informacni_systemy/03-chmelarp.pdf.
- [Chm04] CHMELAŘ, P. *Texturní analýza - metody založené na zpracování signálů* [online]. FIT VUT Brno, 2005. Vedoucí práce Heckel, M. Dostupný z: <https://wis.fit.vutbr.cz/FIT/db/dir.php/rp/2003/RP/1704.pdf>.
- [Chm06] CHMELAŘ, P. *ZZD: Získávání znalostí z multimediálních databází*. FIT VUT Brno, 2006. Vedoucí práce Zendulka, J.
- [COV95] Cortes, C. – Vapnik, V. Support-Vector Networks, *Machine Learning*, 20. 1995.
- [Dar87] WOELK, Darrell – WON, Kim. Multimedia information management in an object-oriented database system. In *Proceedings of the Thirteenth International Conference on Very Large Databases*, Morgan Kaufmann Publishers Inc., 1987...
- [Eis02] EISENBERG, Andrew – MELTON, Jim. *SQL/XML is Making Good Progress* [online]. 2002. [cit. 2006-06-19]. Dostupný z: <http://www.sigmod.org/record/issues/0206/standard.pdf>.
- [Eis04] EISENBERG, Andrew et al. *SQL:2003 Has Been Published* [online]. 2004. [cit. 2006-06-19]. Dostupný z: <http://www.sigmod.org/sigmod/record/issues/0403/E.JimAndrew-standard.pdf>.
- [Eis99] EISENBERG, Andrew – MELTON, Jim. *SQL:1999, formerly known as SQL3* [online]. 1999. [cit. 2006-06-19]. Dostupný z: <http://dbs.uni-leipzig.de/files/lehre/sql99-eisenberg-melton-sigmod-record99.pdf>.
- [Fer06] Fernández-Dívar, I. E. *Image Data Management*. 2006. Vedoucí DP Chmelař, P.
- [Fis06] FISCHER, S. – KLINKENBERG, R. – MIERSWA, I. *YALE 3.3, Yet Another Learning Environment* [online]. 2006 [cit. 2006-09-10]. Dostupný z: http://rapid-i.com/component/option.com_weblinks/catid,24/Itemid,60.
- [Gir06] Giralda, A. R. *Image Databases Indexing*. 2006. Vedoucí MT Zendulka, J.
- [Gra04] Grandvalet, Y. – SANU, C. *Adaptive Scaling for Feature Selection in SVMs* [cit. 2006-09-10]. 2004. Dostupný z: <http://www.hds.utc.fr/~grandval/nips02.pdf>.

- [Han00] HANÁČEK, P. – STAUDEK, J. *Bezpečnost informačních systémů* [online]. 2000 [cit. 2006-09-10]. Dostupný z: <http://www.micr.cz/files/479/uvis-Bezpecnost-20000701.pdf>.
- [Han01] HAN, J., KAMBER, M. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, 2001. ISBN 1-55860-489-8.
- [Hig06] HIGASHI, T. – KITA, S – WATANABE, I. Texture and Objects: Interruption of Same-Object Effect in Human Vision, *IEICE Trans Inf & Syst* E89-D: 1806-1812. 2006.
- [Hin03] HINNER, J. Detekce a rozpoznávání obličejů osob a jejich identifikační význam, *Kriminalistika* [cit. 2006-09-10]. 1/2003. Dostupný z: http://www.mvcr.cz/casopisy/kriminalistika/2003/03_01/hinner.html<http://rfc.net/rfc3261.html>.
- [Hof03] HOFMAN, Jiří. *Formáty multimediálních souborů* [online]. 2006. [cit. 2006-06-19]. Dostupný z: <http://www.aldebaran.cz/studium/formaty.html>.
- [Hru05] HRUŠKA, T. *Informační systémy, Podklady pro studium* [cit. 2006-06-19]. 2005. Dostupný z: <https://www.fit.vutbr.cz/study/courses/INS/private/>.
- [ITU06] International Telecommunication Union. *H.323 : Packet-based multimedia communications systems*. 2006.
- [Jäh99] JÄHNE, Bernd – HAUSSECKER, Horst – GEISLER, Peter. *Handbook of Computer Vision and Applications*. Academic Press, 1999. 3 vol. (624, 942, 894 p.). ISBN 0-12-379770-5.
- [Kad02] KADOUS, M.W. *Dynamic Time Warping* [cit. 2006-09-10]. 2002. Dostupný z: <http://www.cse.unsw.edu.au/~waleed/phd/html/node38.html>.
- [Kol05] KOLÁŘ, D. *Postrelační databáze, Přednášky* [cit. 2006-09-06]. 2005. Dostupný z: <https://www.fit.vutbr.cz/study/courses/PRD/private/>.
- [Kos03] KOSCH, Harald. *Distributed Multimedia Database Technologies Supported by MPEG-7 and MPEG-21*. CRC Press, 2003. 280 s. ISBN 0-8493-1854-8.
- [Kra98] KRÁL, J. *Informační systémy*. Brno: Science, 1998. 357 s. ISBN 80-86083-00-4...
- [Krš05] KRŠEK, Přemysl. *Základy počítačové grafiky, přednášky* [online]. 2005. [cit. 2006-06-19]. Dostupný z: <https://www.fit.vutbr.cz/study/courses/IZG/private/>.
- [Kuc04] KUCHARĚK, T. – KRŮČEK, J. *SQL Multimedia and Application Packages* [online]. 2004 [cit. 2006-09-22]. Dostupný z: http://kocour.ms.mff.cuni.cz/~pokorny/dj/prezentace/2_51.ppt.
- [MP210] ISO/IEC JTC1/SC29/WG11. *MPEG-21 Overview* [online]. Bormans, J. – Hill K. Shanghai: 2002 [cit. 2006-09-14]. Dostupný z: <http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm>.
- [MP402] ISO/IEC JTC1/SC29/WG11. *MPEG-4 Overview*. [online]. Koenen, Rob. 2002 [cit. 2006-05-22]. Dostupný z: <http://www.chiariglione.org/mpeg/standards/mpeg-4/mpeg-4.htm>.
- [MP704] ISO/IEC JTC1/SC29/WG11. *MPEG-7 Overview* [online]. Martínez, José M. Palma de Mallorca : 2004 [cit. 2005-10-22]. Dostupný z: <http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm>.
- [Nos06] NOSEK, F. *Podpora pro práci s multimediálními daty u serveru Oracle* [online]. 2006 [cit. 2006-05-24]. Vedoucí DP Zendulka, J. Dostupný z: <https://wis.fit.vutbr.cz/FIT/db/dir.php/rp/2005/DP/3842.pdf>.
- [ODL05] Oxford Digital Library. *Metadata in the Oxford Digital Library*. [online]. 2005 [cit. 2006-05-16]. Dostupný z: <http://www.odl.ox.ac.uk/metadata.htm>.

- [Or05] Oracle Corporation. *Oracle Database Online Documentation 10g Release2* [online]. 2005 [cit. 2005-10-27]. Dostupný z: http://www.oracle.com/pls/db102/portal.portal_db.
- [Ora06] ORACLE Corporation. *Oracle Database Online Documentation 10g Release2* [online]. 2005 [cit. 2006-05-16]. Dostupný z: http://www.oracle.com/pls/db102/portal.portal_db.
- [Pla04] PLAGEMANN, Thomas et. al. *Using Data Stream Management Systems for Traffic Analysis. A Case Study* [online]. 2004. [cit. 2006-06-19]. Dostupný z: <http://www.pam2004.org/papers/113.pdf>.
- [Raj03] RAJ, C. Pethuru. Multimedia Database, In *Information Technology Portal* [online]. 2003 [cit. 2006-05-25]. Dostupný z: <http://www.peterindia.net/MultimediaDatabase.html>.
- [RCo06] RAD COM. *Voice Over IP Reference Page* [online]. 2006. [cit. 2006-06-19]. Dostupný z: <http://www.protocols.com/pbook/VoIP.htm>.
- [Ros02] ROSENBERG, J., et. al. *SIP: Session Initiation Protocol* [online]. 2002. [cit. 2006-09-10]. Dostupný z: <http://rfc.net/rfc3261.html>.
- [Sor70] SORENSON, H. W. Least-squares estimation: from Gauss to Kalman, *IEEE Spectrum* [online]. 1970. [cit. 2006-09-10] Dostupný z: <http://www.cs.unc.edu/~welch/kalman/media/pdf/Sorenson1970.pdf>.
- [SQL03] Melton, Jim. *SQL (ISO-ANSI Working Draft)*. 2003.
- [Sub98] SUBRAHMANIAN, V. S. *Multimedia database systems*. Morgan Kaufmann, 1998. 442 s. ISBN 1-55860-466-9.
- [Švé05] ŠVÉDA, Miroslav. *PDS: Osnova přednášek* [online]. 2006. [cit. 2006-06-19]. Dostupný z: <https://www.fit.vutbr.cz/study/courses/PDS/private/>.
- [Zbo05] ZBOŘIL, F. *Neuronové sítě, Obsah přednášek* [online]. 2005 [cit. 2006-09-10]. Dostupný z: <https://www.fit.vutbr.cz/study/courses/NEU/private/>.
- [Zem03] ZEMČÍK, Pavel. *Multimédia, ořednášky* [online]. 2003. [cit. 2006-06-19]. Dostupný z: <https://www.fit.vutbr.cz/study/courses/MUM/private/>.
- [Zen05] ZENDULKA, J. *IDS: Materiály k přednáškám ve formátu PDF* [online]. 2005 [cit. 2006-05-16]. Dostupný z: <https://www.fit.vutbr.cz/study/courses/IDS/private/>.

Příloha 1. Rejstřík pojmů

CORBA	(Common Object Request Broker Architecture) definuje API, komunikační protokol, objektový model a služby pro nezávislé, heterogenní a distribuované aplikace
Data	potenciální informace
Databáze	organizovaná kolekce dat
Databázový systém	systém řízení báze dat, více v [Zen05]
DBMS	(DataBase Management System) systém řízení báze dat
Formát média	informace přiřazená médiu v průběhu získávání, komprese nebo přenosu.
Informace	užitečná, použitelná, relevantní nebo zajímavá data
Klíčová slova	ruční popis média
Média	nositel informace, přeneseně multimediální data
Metadata	informace o datech, zde popis (klíčová slova, rysy) a formát média.
MM	MultiMédia
MMDB	MultiMediální DataBáze
Multimédia	digitalizovaná audiovizuální (nestrukturovaná) data; obrázky, audio, video, text
ODMG	(Object Database Management Group) skupina, sjednotila objektový přístup k DB
OLAP	(Online Analytical Processing) analytické funkce, známé jako Business intelligence, data jsou obvykle agregována (sumarizována) do podoby multidimenzionální kostky.
Popis	(description) manuálně nebo automaticky vytvořená charakteristika média, metadata.
Rysy	(features) charakteristika výzazného prvku média
SQL	(Structured Query Language) dotazovací jazyk pro definici, manipulaci, řízení integrity a (souběžného) přístupu k datům na relační bázi
XML	(eXtensible Markup Language) univerzální značkovací jazyk
ISO	(International Organization for Standardization) mezinárodní normalizační institut (v ČR zastoupen ČNI – Český normalizační institut), standard, norma (obdoba ČSN)
VoIP	(Voice over IP) hlas (a v současné době i video) přenášené protokolem IP
CBR	(Content Based Retrieval) vyhledávání dle obsahu dat.

Další bude doplněno

Příloha 2. Oracle interMedia aplikace

Pro tvorbu libovolné aplikace je stěžejní výběr prostředků. Databází zde bude Oracle Database 10g včetně rozšíření interMedia. Pro přístup Oracle doporučuje objektově-relační přístup pomocí jazyků Java a PL/SQL i když je k multimediálním datům možné přistupovat i prostřednictvím relačního modelu a tedy téměř libovolného procedurálního strukturovaného jazyka (pro seznámení s alternativami začněte na: www.oracle.com/technology/sample_code/products/intermedia/).

Pro následující ukázkou připojení, uploadu a downloadu obrázku bude použita Java a Oracle JDeveloper, který umožňuje vývoj okenních i webových aplikací.

Připojení k databázi

Nejprve je nutné se připojit se k databázi a potřebné funkce importovat:

```
// JDBC a SQLJ
import java.sql.Connection;
import java.sql.SQLException;
import java.sql.DriverManager;
import java.sql.Statement;
import java.sql.PreparedStatement;
import java.sql.ResultSet;
import oracle.jdbc.OracleResultSet;
import oracle.jdbc.OraclePreparedStatement;
// set the CLASSPATH variable
set CLASSPATH = %ORACLE_HOME%\jdbc\lib\ojdbc14.jar;
                %ORACLE_HOME%\sqlj\lib\runtime12.jar;

// interMedia
import oracle.ord.im.OrdImage;
import oracle.ord.im.OrdImageSignature;
// set the CLASSPATH variable
set CLASSPATH=%ORACLE_HOME%\ord\jlib\ordim.jar;
```

Potřebné funkce pro práci s JDBC a MM daty jsou obsažené v balíčcích dodávaných spolu s JDeveloperem (v adresáři %ORACLE_HOME%, ve škole Q:\ORACLE\JDeveloper).

Pro připojení je nutné registrovat ovladač, připojit se k databázi – na serveru **pcuifs1** v současné době běží Oracle Database 10g. Jako jméno a heslo je obvykle přednastavený xlogin. Pro práci s velkými binárními daty (dvojfázový uzamykací protokol) je nutné provádět **commit** ručně(!) a inicializovat SQL statement:

```
// register the oracle JDBC driver with the driver manager
DriverManager.registerDriver(new oracle.jdbc.driver.OracleDriver());

// connect pcuifs1.fit.vutbr.cz
Connection conn = DriverManager.getConnection("jdbc:oracle:thin:"
+ "login/password"
+ "@pcuifs1.fit.vutbr.cz:1521:stud");

// Note: it is CRITICAL to set the autocommit to false
conn.setAutoCommit(false);

// create a JDBC Statement object to execute SQL in the database
Statement stmt = conn.createStatement();
```

Základní datové typy

Multimediální data je možné v databázi uložit jako BLOB, ale výhodnější je využití speciálních UDT. V případě obrázku je to typ **ORDSYS.ORDImage**, jako na Fig. 5.1.

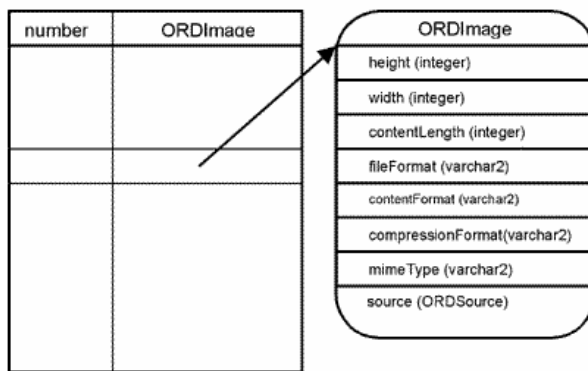


Fig. 5.1 Struktura typu ORDImage v databázi Oracle [Or05]

Pro přístup k databázi se v Javě využívá JDBC, které obecně nepodporuje objektový přístup. Proto jsou vyvinuty Java objekty pro manipulaci s těmito daty, nicméně všechny operace v objektově-relační databázi nadále vykonává DBMS, proto musí být vytvořeny jako proxy objekty (**ORDImage**) z databázového objektu, jak je uvedeno na obrázku níže.

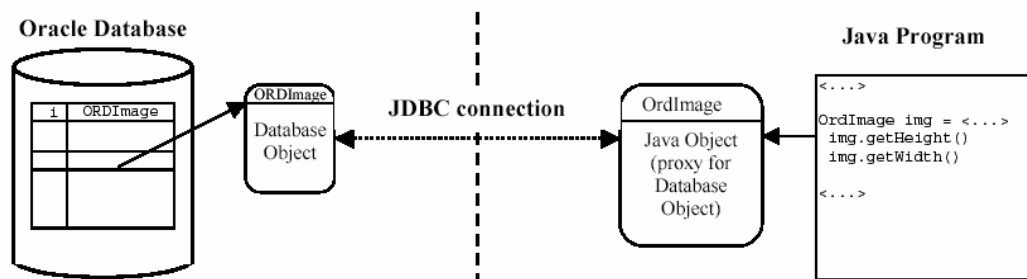


Fig. 5.2 Schéma použití ORDImage jako objektu v databázi a proxy v Javě [Or05]

Příklad vytvoření tabulky pro uchovávání obrázků:

```
CREATE TABLE image_table
(id number primary key,
 image ordsys.ordimage,
 signature ordsys.ordimagesignature
);

CREATE SEQUENCE image_sequence
INCREMENT BY 1 START WITH 1;
```

Vytvořená tabulka obsahuje sloupec pro uchování obrázku (**image**) a popis tohoto obrázku (**signature**) pro následující vyhledávání dle obsahu a sloupec (**id**) jako primární klíč, jehož unikátní hodnoty se získávají ze sekvence definované pod ní.

Vložení obrázku

Nejprve načteme hodnotu primárního klíče ze sekvence (**nextval**) a poté jej uložíme spolu s inicializovanými objekty obrázku a jeho popisu.

```
// retrieve the next value in the sequence order
String SQLQuery = ("SELECT image_sequence.nextval FROM dual");
OracleResultSet rset = (OracleResultSet)stmt.executeQuery(SQLQuery);
rset.next();
int nextval = rset.getInt("nextval");

// insert a row (nextval, initialized ORDImage) into image_table
SQLQuery = "INSERT INTO image_table (id, image, signature) VALUES (" +
nextval + ", ordsys.ordimage.init(), ordsys.ordimagesignature.init())";
```

```
stmt.execute(SQLquery);
```

Protože Javové proxy objekty musí být vytvořeny z databázových objektů, musíme právě vložené objekty pro obrázek a jeho popis načíst pro další úpravu:

```
// select the new ORDImage database object for update
SQLquery = "SELECT image, signature FROM image_table WHERE id = "
           + nextval + " FOR UPDATE";
rset = (OracleResultSet)stmt.executeQuery(SQLquery);
rset.next();

// select the image into a java proxy OrdImage object
OrdImage imageProxy =
    (OrdImage)rset.getORAData("image", OrdImage.getORADataFactory());
// if not working use deprecated version of JDBC/SQLJ:
// (OrdImage)rset.getCustomDatum("image", OrdImage.getFactory());

// select the signature into a java proxy OrdImageSignature object
OrdImageSignature signatureProxy =
    (OrdImageSignature)rset.getCustomDatum("signature",
                                           OrdImageSignature.getFactory());
rset.close();
```

Až teprve teď máme zinicizované objekty a obrázek se může fyzicky načíst:

```
// load the image data from file "fileName" into the ORDImage object
imageProxy.loadDataFromFile(fileName);
```

Předchozí funkce nad proxy objektem způsobí, že se obrázek načte nejen do lokálního objektu, ale také do databáze, která poté provádí všechny operace nad ním. Jsou to především operace pro inicializaci metadat (vlastností) a popis obsahu:

```
// automatically detect the image's height, width, file format...
// execute ORDImage.setProperties() on the server
imageProxy.setProperties();

// generate an image signature for the specified image on server
signatureProxy.generateSignature(imageProxy);
```

Nyní je nutné provést update a commit, aby byly provedené operace reflektovány:

```
SQLquery = "UPDATE image_table SET image=?, signature=? where id=?";
OraclePreparedStatement opstmt =
    OraclePreparedStatement(conn.prepareStatement(SQLquery));
opstmt.setCustomDatum(1, imageProxy);
opstmt.setCustomDatum(2, signatureProxy);
opstmt.setInt(3, nextval);
opstmt.execute();
opstmt.close();

// commit all changes
conn.commit();
```

V textové reprezentaci SQL dotazu je poněkud obtížné zapsat (referenci na) obrázek a jeho popis (to lze například pomocí MIME), proto je vhodné doraz předpřipravit a poté do něj vložit uživatelská data, nahrazením (??) funkcí `setCustomDatum`.

Dotazování vlastností a manipulace s obrázkem

Obrázek je nyní perzistentně uložen v databázi a je možné jej dotazovat. Nejprve si ukážeme, jak lze zjistit velikost a další vlastnosti:

```
// select the image into the proxy object (imageProxy)
String SQLquery = "select image from image_table where id = 1";
OracleResultSet rset = (OracleResultSet)stmt.executeQuery(SQLquery);
rset.next();
OrdImage imageProxy =
    (OrdImage)rset.getORADData("image", OrdImage.getORADDataFactory());
// (OrdImage)rset.getCustomDatum("image", OrdImage.getFactory());
rset.close();

// and call getXY methods ... type imageProxy. & press CTRL+SPACE
int height = imageProxy.getHeight();
int width = imageProxy.getWidth();
```

Poté, pokud například chceme uložit jeho náhled, provedeme následující zpracování:

```
SQLquery = ("insert into image_table (id, image) " +
            "values (2, ordsys.ordimage.init())");
...
// get the newly inserted destination OrdImage object
SQLquery = "select image from image_table where id=2 for update";
...
OrdImage dstImageProxy =
    (OrdImage)rset.getORADData("image", OrdImage.getORADDataFactory());

// call the processCopy method (processing occurs on the SERVER)
ImageProxy.processCopy(
    "maxscale=100 100 fileformat=jfif", dstImageProxy );

// update the destination image in the second row
String dstUpdateSQL = "update image_table set image=? where id=2";
opstmt.setORADData(1, dstImageProxy);
// commit
```

Uvedená funkce **processCopy** zajišťuje kopii po zpracování obrázku se zadanými parametry (viz JavaDoc [Or05]), obdobně samotné zpracování funkcí **process**.

Stažení obrázků a jejich uložení do souboru je také poměrně jednoduchá operace:

```
// export the data in row 2
String SQLquery = "select image from image_table where id = 2";
OracleResultSet rset = (OracleResultSet)stmt.executeQuery(SQLquery);
rset.next();
OrdImage imageProxy =
    (OrdImage)rset.getORADData("image", OrdImage.getORADDataFactory());
rset.close();

// call the getDataInFile method to write the image in row 2 to disk
imageProxy.getDataInFile("row2.jpg");
```

Dotazování jiných typů multimedií (zvuk, video, dokumenty) je obdobné, jen obsahují jiné funkce pro jejich manipulaci (v Javě i databázi):

```
import oracle.ord.im.OrdAudio;
import oracle.ord.im.OrdDoc;
import oracle.ord.im.OrdImage;
import oracle.ord.im.OrdVideo;

String SQLquery = "select product_photo, " +
    " product_audio," +
    " product_video, " +
    " product_documents from products where product_id=3117";
PreparedStatement pstmt = conn.prepareStatement(SQLquery);
```

```
OracleResultSet rset = (OracleResultSet)pstmt.executeQuery();
if ( rset.next() )
{
    OrdImage imgProxy = (OrdImage)rset.getORADData(
        "product_photo", OrdImage.getORADDataFactory());
    OrdAudio audProxy = (OrdAudio)rset.getORADData(
        "product_audio", OrdAudio.getORADDataFactory());
    OrdVideo vidProxy = (OrdVideo)rset.getORADData(
        "product_video", OrdVideo.getORADDataFactory());
    OrdDoc docProxy = (OrdDoc)rset.getORADData(
        "product_testimonials", OrdDoc.getORADDataFactory());
}

String audFormat = audProxy.getFormat();
String vidMimeType = vidProxy.getMimeType();
```

Mazání je provedeno obvyklým způsobem, pomocí SQL:

```
DELETE FROM image_table

// drop image table and sequence
stmt.executeQuery("DROP TABLE image_table");
stmt.executeQuery("DROP SEQUENCE image_sequence");
// commit all changes
conn.commit();
```

Vyhledávání dle podobnosti

Vyhledávání dle obsahu je jedna z věcí, které odlišují Oracle od běžných databázových producentů. Před vlastním vyhledáváním je nutné vytvořit tento popis dle obsahu (**ORDImageSignature**) pomocí funkce **generateSignature**. Vlastní dotaz je možné provést užitím spárovaných funkcí (123) pomocí SQL:

```
SELECT img1.id, img2.id, ORDSYS.IMGScore(123) as distance
FROM image_table img1, image_table img2
WHERE ORDSYS.IMGSimilar(img1.signature, img2.signature, 'color="0.3"
texture="0.3" shape="0.3" location="0.1"', 50, 123) = 1
AND img1.id <> img2.id
AND img1.id = 1
ORDER BY distance
```

K vlastním funkcím: **IMGSimilar** porovnává dvě signatury a vrací hodnotu (0 / 1), která indikuje, zda jsou obrázky dle váhy zadaných rysů (barva, textura, tvar a umístění) podobné více, než specifikovaný treshold (50). Spřátelená funkce **IMGScore** vrací míru podobnosti těchto obrázků, kde 0 znamená totožnost a 100 naprostou odlišnost.